

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 August 2004 (12.08.2004)

PCT

(10) International Publication Number
WO 2004/067706 A2

(51) International Patent Classification⁷:

C12N

70810 (US). **FIORETTI, William, C.** [US/US]; 2225 Lak-
eridge Road Drive, Grapevine, TX 76051 (US). **CADD,**
Gary, G. [US/US]; 501 Turner Road, Apartment 1111,
Grapevine, TX 76051 (US).

(21) International Application Number:

PCT/US2003/041261

(22) International Filing Date:

24 December 2003 (24.12.2003)

(74) Agent: **PRATT, John, S.**; Kilpatrick Stockton, LLC, Suite
2800, 1100 Peachtree Street, Atlanta, GA 30309 (US).

(25) Filing Language:

English

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR,
CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD,
GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN,
MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU,
SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(26) Publication Language:

English

(84) Designated States (regional): ARIPO patent (BW, GH,
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE,
SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA,
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(30) Priority Data:

60/441,377	21 January 2003 (21.01.2003)	US
60/441,381	21 January 2003 (21.01.2003)	US
60/441,392	21 January 2003 (21.01.2003)	US
60/441,405	21 January 2003 (21.01.2003)	US
60/441,447	21 January 2003 (21.01.2003)	US
60/441,502	21 January 2003 (21.01.2003)	US
10/609,019	26 June 2003 (26.06.2003)	US

(71) Applicants (for all designated States except US): **TRANS-GENRX, INC.** [US/US]; 1750 Valley View Lane, Suite 110, Dallas, TX 75243 (US). **THE BOARD OF SUPERVISORS OF LOUISIANA STATE UNIVERSITY AND AGRICULTURAL AND MECHANICAL COLLEGE** [US/US]; P.O. box 25055, Baton Rouge, LA 70894 (US).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and

(75) Inventors/Applicants (for US only): **COOPER, Richard, K.** [US/US]; 111 Pecan Meadow Drive, Baton Rouge, LA

A2

(54) Title: PRODUCTION OF MULTIMERIC PROTEINS

Prom	prepro	Gene of Interest	pro	Gene of interest	polyA
------	--------	------------------	-----	------------------	-------

(57) **Abstract:** The present invention provides a new, effective and efficient method of producing multimeric proteins in an individual. Multimeric proteins include associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single polypeptide encoded by more than one gene of interest). Expression and/or formation of the multimeric protein in the individual is achieved by administering a polynucleotide cassette containing genes of interest that encode portions of the multimeric protein to the individual. The polynucleotide cassette may additionally contain one or more pro sequences, prepro sequences, cecropin prepro sequences, and/or cleavage site sequences.

WO 2004/067706

PRODUCTION OF MULTIMERIC PROTEINS

The U.S. Government has certain rights in this invention. The development of this invention was partially funded by the United States Government under a HATCH 15 grant from the United States Department of Agriculture, partially funded by the United States Government with Formula 1433 funds from the United States Department of Agriculture and partially funded by the United States Government under contract DAAD 19-02016 awarded by the Army.

FIELD OF THE INVENTION

The present invention relates generally to production of multimeric proteins in a transgenic individual, wherein genes encoding the multimeric proteins are operably-linked to signal sequences, or portions of signal sequences.

BACKGROUND OF THE INVENTION

Methods for producing multimeric proteins in transgenic animals are desirable for a variety of reasons, including the transgenic animal's potential as biological factories to produce multimeric proteins for pharmaceutical, diagnostic and 30 industrial uses. This potential is attractive to the industry due to the inadequate capacity in facilities used for recombinant production of multimeric proteins and the increasing demand by the pharmaceutical industry for use of these facilities. Numerous attempts to produce transgenic animals have met several problems, including low rates of gene incorporation and unstable gene incorporation. 35 Accordingly, improved gene technologies are needed for the development of transgenic animals for the production of multimeric proteins.

Several of the prior art gene delivery technologies employed viruses that are associated with potentially undesirable side effects and safety concerns. The majority

of current gene-delivery technologies useful for gene therapy rely on virus-based delivery vectors, such as adeno and adeno-associated viruses, retroviruses, and other viruses, which have been attenuated to no longer replicate. (Kay, M.A., et al. 2001. *Nature Medicine* 7:33-40).

5 There are multiple problems associated with the use of viral vectors. Firstly, they are not tissue-specific. In fact, a gene therapy trial using adenovirus was recently halted because the vector was present in the patient's sperm (Gene trial to proceed despite fears that therapy could change child's genetic makeup. *The New York Times*, December 23, 2001). Secondly, viral vectors are likely to be transiently
10 incorporated, which necessitates re-treating a patient at specified time intervals. (Kay, M.A., et al. 2001. *Nature Medicine* 7:33-40). Thirdly, there is a concern that a viral-based vector could revert to its virulent form and cause disease. Fourthly, viral-based vectors require a dividing cell for stable integration. Fifthly, viral-based vectors indiscriminately integrate into various cells, which can result in undesirable germline
15 integration. Sixthly, the required high titers needed to achieve the desired effect have resulted in the death of one patient and they are believed to be responsible for induction of cancer in a separate study. (Science, News of the Week, October 4, 2002).

20 Accordingly, what is needed is a new method to produce multimeric proteins in transgenic animals and humans, in which the vector containing those genes does not cause disease or other unwanted side effects. There is also a need for DNA constructs that would be stably incorporated into the tissues and cells of animals and humans, including cells in the resting state that are not replicating. There is a further
25 recognized need in the art for DNA constructs capable of delivering genes to specific tissues and cells of animals and humans and for producing multimeric proteins in those animals and humans.

SUMMARY OF THE INVENTION

30 The present invention provides a new, effective and efficient method of producing multimeric proteins in an individual. Multimeric proteins include associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single polypeptide encoded by more than one gene of interest). Expression and/or formation of the multimeric protein in the individual is achieved by
35 administering a polynucleotide cassette containing the genes of interest to the individual. The polynucleotide cassette may additionally contain one or more pro sequences, prepro sequences, cecropin prepro sequences, and/or cleavage site sequences.

This invention provides polynucleotide cassettes containing two or more genes of interest and two or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. Each of the genes of interest encodes a polypeptide that forms a part of the multimeric protein. One discovery of 5 the present invention is the use of pro portions of prepro signal sequences to facilitate appropriate processing, expression, and/or formation of multimeric proteins in an individual. Several examples of prepro polynucleotides from which a pro polynucleotide can be derived or be a part of are a cecropin prepro, lysozyme prepro, ovomucin prepro, ovotransferrin prepro, a signal peptide for tumor necrosis factor 10 receptor (SEQ ID NO:6), a signal peptide encoded by a polynucleotide sequence provided in one of SEQ ID NOs:7-54 and a signal peptide provide in SEQ ID NO:55. The prepro or pro polynucleotide can be a cecropin prepro or pro polynucleotide selected from the group consisting of cecropin A1, cecropin A2, cecropin B, cecropin C, cecropin D, cecropin E and cecropin F. In a preferred embodiment, the pro 15 polynucleotide is a cecropin B pro polynucleotide having a sequence shown in SEQ ID NO:1 or SEQ ID NO:2. A preferred prepro polynucleotide is a cecropin B polynucleotide having a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

Another discovery of the present invention is that cecropin prepro sequences facilitate appropriate processing, expression, and/or formation of proteins, including 20 multimeric proteins, in an individual. Accordingly, the present invention includes polynucleotide cassettes containing one or more genes of interest operably-linked to a cecropin prepro sequence. In one embodiment, the polynucleotide cassette contains two or more genes of interest operably-linked to a cecropin prepro sequence. Preferred cecropin prepro polynucleotides are provided in SEQ ID NO:3 and SEQ ID 25 NO:4. The present invention also includes polynucleotide cassettes containing two or more genes of interest operably linked to a cecropin prepro polynucleotide, wherein pro sequences are located between the genes of interest.

These polynucleotide cassettes are administered to an individual for expression of polypeptide sequences and the formation of a protein, and more 30 preferably, a multimeric protein. Preferably, the individual is an animal from which the protein can be harvested. Preferred animals are egg-laying or milk-producing animals.

In one embodiment, the egg-laying transgenic animal is an avian. The method of the present invention may be used in avians including Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones.

5 Preferably, the egg-laying transgenic animal is a poultry bird. More preferably, the bird is a chicken, turkey, duck, goose or quail. Another preferred bird is a ratite, such as, an emu, an ostrich, a rhea, or a cassowary. Other preferred birds are partridge, pheasant, kiwi, parrot, parakeet, macaw, falcon, eagle, hawk, pigeon, cockatoo, song birds, jay bird, blackbird, finch, warbler, canary, toucan, mynah, or sparrow.

10 In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual. The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding 15 sequence for the protein or peptide of interest and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the 20 codons for the first several N-terminal amino acids of the transposase, wherein the nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of 25 the transposase gene. In some embodiments, the effective polyA sequence is an avian optimized polyA sequence.

In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the 30 pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino

acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance 5 expression of the transposase gene. One example of such a transposon-based vector is the pTnMod vector (SEQ ID NO:57).

Accordingly, it is an object of the present invention to provide improved methods for the production of multimeric proteins in an individual.

It is another object of the present invention to provide improved methods for 10 the production of multimeric proteins in an egg-laying animal or a milk-producing animal.

It is yet another object of the present invention to provide improved methods for the production of multimeric proteins in a chicken or quail.

Another object of the present invention is to provide a method to produce an 15 egg or milk containing a multimeric protein.

An advantage of the present invention is that multimeric proteins are produced by transgenic animals much more efficiently and economically than prior art methods, thereby providing a means for large scale production of multimeric proteins.

These and other objects, features and advantages of the present invention will 20 become apparent after a review of the following detailed description of the disclosed embodiments and claims.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 depicts schematically a polynucleotide cassette containing two genes 25 of interest operably-linked to two pro polynucleotides, wherein the first pro polynucleotide is a part of a prepro polynucleotide. "Prom" indicates promoter.

Figure 2 depicts schematically a polynucleotide cassette containing 30 polynucleotides encoding for a heavy chain and a light chain of an antibody. "Oval prom" indicates an ovalbumin promoter. The polynucleotide cassette contains pro and prepro sequences and is flanked by insertion sequences (IS) recognized by a transposase.

Figure 3 depicts schematically a polynucleotide cassette containing a cecropin prepro sequence operably-linked to two genes of interest. Between the genes of interest resides a cleavage site indicated by "CS."

5 Figure 4 depicts schematically a polynucleotide cassette containing two genes of interest, a promoter (prom), a signal sequence (SS) and a cleavage site (CS). The polynucleotide cassette is flanked by insertion sequences (IS) recognized by a transposase.

10 Figure 5 is a picture of a gel showing partially purified egg white derived from a transgenic avian run under reducing and non-reducing conditions.

DETAILED DESCRIPTION OF THE INVENTION

15 The present invention provides a new, effective and efficient method of producing multimeric proteins in an individual. Multimeric proteins include associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single polypeptide encoded by more than one gene of interest). Expression and/or formation of the multimeric protein in the individual is achieved by administering a polynucleotide cassette containing the genes of interest to the 20 individual. The polynucleotide cassette may additionally contain one or more pro sequences, prepro sequences, cecropin prepro sequences, and/or cleavage site sequences.

25 This invention provides polynucleotide cassettes containing two or more genes of interest and two or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. Each of the genes of interest encodes a polypeptide that forms a part of the multimeric protein. These polynucleotide cassettes are administered to an individual for expression of the polypeptide sequences and expression and/or formation of the multimeric protein. Preferably, the individual is an animal from which the multimeric protein can be 30 harvested. Preferred animals are egg-laying or milk-producing animals. In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual.

The pro polynucleotide sequences operably-linked to the genes of interest include pro portions of prepro polynucleotide sequences commonly associated with

polynucleotides encoding proteins secreted from a cell in nature. It may be that the pre polynucleotide sequence functions to direct the resultant protein into the endoplasmic reticulum and the pro sequence is cleaved within the endoplasmic reticulum or Golgi complex of a cell containing the protein. While prepro 5 polynucleotide sequences are associated with secreted polypeptides in nature, one discovery of the present invention is the use of pro portions of the prepro signal sequences to facilitate appropriate processing, expression, and/or formation of multimeric proteins, and more particularly, associated multimeric proteins. In the present invention, each gene of interest is operably-linked with a pro polynucleotide 10 sequence. Figure 1 shows schematically one polynucleotide cassette containing two genes of interest, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The first gene of interest is operably-linked to a pro polynucleotide sequence that is part of a prepro polynucleotide sequence, while the second gene of interest is operably-linked to a pro polynucleotide sequence that is not 15 part of a prepro polynucleotide sequence, but may have been derived from a prepro polynucleotide sequence. Accordingly, the term "pro sequence" encompasses a pro sequence that is part of a prepro sequence and a pro sequence that is not part of a prepro sequence, but may have been derived from a prepro sequence. In preferred embodiments, the most 5' pro polynucleotide sequence in the polynucleotide cassette 20 is a part of a prepro polynucleotide sequence.

Several examples of prepro polynucleotides from which a pro polynucleotide can be derived or be a part of are a cecropin prepro, lysozyme prepro, ovomucin prepro, ovotransferrin prepro, a signal peptide for tumor necrosis factor receptor (SEQ ID NO:6), a signal peptide encoded by a polynucleotide sequence provided in one of 25 SEQ ID NOS:7-54 and a signal peptide provided in SEQ ID NO:55. The prepro or pro polynucleotide can be a cecropin prepro or pro polynucleotide selected from the group consisting of cecropin A1, cecropin A2, cecropin B, cecropin C, cecropin D, cecropin E and cecropin F. In a preferred embodiment, the pro polynucleotide is a cecropin B pro polynucleotide having a sequence shown in SEQ ID NO:1 or SEQ ID NO:2. A 30 preferred prepro polynucleotide is a cecropin B polynucleotide having a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

Figure 1 provides one embodiment of the invention wherein the polynucleotide cassette includes two genes of interest and two pro polynucleotide

sequences arranged in the following order: a prepro polynucleotide, a first gene of interest, a pro polynucleotide, and a second gene of interest. Preferably, the sequences are arranged in the aforementioned order beginning at a 5' end of the polynucleotide cassette. Figure 2 provides a more specific embodiment of the present invention wherein the first and second genes of interest are polynucleotides encoding antibody heavy and light chains. However, the invention includes polynucleotide cassettes containing at least two genes of interest. Each of the genes of interest is operably-linked to a pro polynucleotide. Each of these pro polynucleotides can be the same, or each can be different. In one embodiment, all of the pro polynucleotides in the polynucleotide cassette are the same and are cecropin pro polynucleotides. The most 5' cecropin pro polynucleotide is preferably a part of a cecropin prepro polynucleotide sequence as shown in Figure 3.

The polynucleotide cassettes of the present invention may be administered to an individual for production of a multimeric protein in that individual. Accordingly, the present invention includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising at least two genes of interest, each encoding a part of the multimeric protein, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The present invention also includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising a cecropin prepro sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein. This second method does not require the linking of pro polynucleotides to each gene of interest since the use of a cecropin prepro sequence itself in a polynucleotide cassette facilitates processing, expression, and/or formation of multimeric proteins. Polynucleotide cassettes containing the cecropin prepro polynucleotide can contain at least two genes of interest. Preferably, the cecropin prepro polynucleotide is located 5' of the genes of interest in the polynucleotide cassette. One exemplary polynucleotide cassette is shown in Figure 3. In a preferred embodiment, the prepro sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4. As shown in Figure 3, the polynucleotide cassettes containing a cecropin prepro polynucleotide preferably contain a cleavage site between each of two genes of interest. Such cleavage site(s) may be nucleotides encoding any cleavage sites including, but not

limited to, an enzymatic cleavage site, a pro polynucleotide, and a photolabile cleavage site, a chemical cleavage site, and a self-splicing cleavage site (i.e., intein). Cleavage sites are discussed in more detail below.

The polynucleotide cassettes of the present invention are particularly suited for 5 production of multimeric proteins in an individual. Individuals include both humans and animals. Preferred animals are egg-laying animals and milk-producing animals. As used herein, the term "egg-laying animal" includes all amniotes such as birds, turtles, lizards and monotremes. Monotremes are egg-laying mammals and include the platypus and echidna. The term "bird" or "fowl," as used herein, is defined as a 10 member of the Aves class of animals which are characterized as warm-blooded, egg-laying vertebrates primarily adapted for flying. Avians include, without limitation, Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. The term "Ratite," as used herein, is defined as a 15 group of flightless, mostly large, running birds comprising several orders and including the emus, ostriches, kiwis, and cassowaries. The term "Psittaciformes", as used herein, includes parrots and refers to a monofamilial order of birds that exhibit zygodactylism and have a strong hooked bill. A "parrot" is defined as any member of the avian family Psittacidae (the single family of the Psittaciformes), distinguished by 20 the short, stout, strongly hooked beak. Preferred avians are poultry birds including chickens, quail, turkeys, geese and ducks. The term "chicken" as used herein denotes chickens used for table egg production, such as egg-type chickens, chickens reared for public meat consumption, or broilers, and chickens reared for both egg and meat production ("dual-purpose" chickens). The term "chicken" also denotes chickens 25 produced by primary breeder companies, or chickens that are the parents, grandparents, great-grandparents, etc. of those chickens reared for public table egg, meat, or table egg and meat consumption.

When the polynucleotide cassettes of the present invention are administered to 30 an egg-laying or milk-producing animal, a transgenic animal containing a polynucleotide cassette is created and the animal produces a transgenic multimeric protein. It is preferred that the resultant multimeric protein is deposited in the egg or in the milk. Various different signal sequences and promoters may be used to achieve deposition of the multimeric protein in the egg or in the milk and these are described

in more detail below. In order to achieve a transgenic animal containing a polynucleotide cassette of the present invention, the polynucleotide cassettes can be administered to the individual with, or contained in, any vector, as naked DNA, or in any delivery construct or solution. A preferred vector for incorporation of the 5 polynucleotide cassettes into an individual is a transposon-based vector described below.

Definitions

It is to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for 10 example, reference to "a cell" can mean that at least one cell can be utilized.

The term "antibody" is used interchangeably with the term "immunoglobulin" and is defined herein as a protein synthesized by an animal or a cell of the immune system in response to the presence of a foreign substance commonly referred to as an "antigen" or an "immunogen". The term antibody includes fragments of antibodies. 15 Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to as an "antigenic determinant" or an "epitope". Antigens can be naturally occurring or artificially engineered. Artificially engineered antigens include but are not limited to small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. 20 Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

25 The term "egg" is defined herein as including a large female sex cell enclosed in a porous, calcarous or leathery shell, produced by birds and reptiles. The term "ovum" is defined as a female gamete, and is also known as an egg. Therefore, egg production in all animals other than birds and reptiles, as used herein, is defined as the production and discharge of an ovum from an ovary, or "ovulation". Accordingly, it 30 is to be understood that the term "egg" as used herein is defined as a large female sex cell enclosed in a porous, calcarous or leathery shell, when a bird or reptile produces it, or it is an ovum when it is produced by all other animals.

The term "gene" is defined herein to include a polynucleotide that includes a coding region for a protein, peptide or polypeptide, with or without intervening sequences such as introns.

The term "multimeric protein" is defined herein to include one or more polypeptides that are associated, or joined, by any means including disulfide bonds. An example of this type of multimeric protein is an antibody that contains both heavy and light chains that are associated by disulfide bonds. These multimeric proteins are referred to herein as "associated multimeric proteins." The term "multimeric protein" also includes a polypeptide that is encoded by more than one gene of interest. An example of this type of multimeric protein is a single polypeptide containing a heavy chain polypeptide (first polypeptide of interest) and a light chain polypeptide (second polypeptide of interest). In these embodiments, the different polypeptides of interest may be separated by other polypeptide sequences such as spacer polypeptides and cleavage site polypeptides. These types of multimeric proteins are referred to herein as "multivalent multimeric proteins."

The term "milk-producing animal" refers herein to mammals including, but not limited to, bovine, ovine, porcine, equine, and primate animals. Milk-producing animals include but are not limited to cows, llamas, camels, goats, reindeer, zebu, water buffalo, yak, horses, pigs, rabbits, non-human primates, and humans.

The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA incorporated into its DNA. While a transgenic animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more somatic cells that contain a portion of the transposon-based vector DNA for any period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into the animal's DNA for a period of at least five days, more preferably the laying life of the animal, and most preferably the life of the animal. In a further preferred embodiment, the animal is an avian.

The term "vector" is used interchangeably with the terms "construct", "DNA construct" and "genetic construct" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or

organisms, such as birds. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

When referring to two nucleotide sequences, one being a regulatory sequence, 5 the term "operably-linked" is defined herein to mean that the two sequences are associated in a manner that allows the regulatory sequence to affect expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

The term "regulatory sequence" is defined herein as including promoters, 10 enhancers and other expression control elements such as polyadenylation sequences, matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance expression, and silencers.

Transposon-Based Vectors

15 While not wanting to be bound by the following statement, it is believed that the nature of the DNA construct is an important factor in successfully producing transgenic animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species, especially avians, have low efficiencies and may constitute a major reason for the low rates of 20 transformation previously observed. The DNA (or RNA) constructs previously used often do not integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to 25 an animal that overcome the prior art problems relating to low transgene integration frequencies. Two preferred transposon-based vectors of the present invention in which a transposase, gene of interest and other polynucleotide sequences may be introduced are termed pTnMCS (SEQ ID NO:56) and pTnMod (SEQ ID NO:57).

The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous 30 vectors. More specifically, intratesticular injections performed with a prior art transposon-based vector (described in U.S. Patent No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive

roosters. Actual frequencies of integration were estimated by either or both comparative strength of the PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of one or more of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is the pTnMod vector (SEQ ID NO:57). The transposon-based vector may additionally or alternatively include one or more of the following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:58), AAGATGT (SEQ ID NO:59), ACGATGA (SEQ ID NO:60),

AAGATGG (SEQ ID NO:61), GACATGA (SEQ ID NO:62), ACCATGA (SEQ ID NO:63), and ACCATGA (SEQ ID NO:64), ACCATGT (SEQ ID NO:65).

Transposases and Insertion Sequences

In a further embodiment of the present invention, the transposase found in the 5 transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, 10 Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 ($\gamma\delta$), Tn1681, Tn2901, AC transposons, Mp transposons, Spm transposons, En transposons, Dotted transposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more 15 modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the promoter operably-linked to the transposase; b) a change of one or more of the codons for the first several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase 20 synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene facilitate transcription of the transposase gene, in part, by increasing strand 25 dissociation during transcription. It is preferable that one or more of between approximately the first 1 to 20, more preferably 3 to 15, and most preferably between 4 to 12 N-terminal codons of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment, the first ten N-terminal codons of the transposase gene are modified 30 in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Patent No. 5,719,055.

In some embodiments, the transposon-based vectors are optimized for expression in a particular host by changing the methylation patterns of the vector DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

Transposases and insertion sequences from other analogous eukaryotic transposon-based vectors that can also be modified and used are, for example, the Drosophila P element derived vectors disclosed in U.S. Patent No. 6,291,243; the Drosophila mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA, 98:6759-6764; L. Zagoraiou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), Mobile DNA, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

Many transposases recognize different insertion sequences, and therefore, it is to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the transposase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type transposons that typically contain insertion sequences of well over 100 base pairs.

While the examples provided below incorporate a “cut and insert” Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a “rolling replication” type transposon-based vector. Use of a rolling replication type transposon allows multiple copies of the transposon/transgene to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use of a rolling replication system may require tight control to limit the insertion events to

non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

Stop Codons and PolyA Sequences

5 In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest. While not wanting to be bound by the following statement, it is thought that the stop codon UAG is less effective in translation
10 termination and is therefore less desirable in the constructs described herein.

As used herein an “effective polyA sequence” refers to either a synthetic or non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A polyA sequence may be operably-linked to any gene
15 in the transposon-based vector including, but not limited to, a transposase gene and a gene of interest. A preferred polyA sequence is optimized for use in the host animal or human. In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40
20 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string and thereby separate the stop codon from the A polynucleotide string. In one embodiment of the present invention, the polyA sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:66 and as taken from GenBank accession # Y00407, base pairs 10651-11058. In another
25 embodiment, the polyA sequence comprises a synthetic polynucleotide sequence shown in SEQ ID NO:67. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO:68. A chicken optimized polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

30 It is a surprising discovery of the present invention that such an avian optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. Accordingly, the present invention includes methods of or increasing incorporation of

a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a 5 gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A 10 polynucleotide string. The present invention further provides transposon-based vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

Promoters and Enhancers

The first promoter operably-linked to the transposase gene and the second 15 promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters, β -actin promoter, tubulin promoter, Rous-Sarcoma 20 virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters and chemically inducible promoters. Examples of tissue-specific promoters include the glucose 6 phosphate (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin 25 promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a polynucleotide sequence of SEQ ID NO:69. The G6P promoter sequence may be deduced from a rat G6P gene untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters 30 and several hormone induced promoters. Examples of chemically inducible promoters include reproductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible metallothionein promoter.

Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. Genes and Development, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. Mol. Ther. 1:159-164); estrogen-based inducible systems (Braselmann, S. et al. 1993. Proc. Natl. Acad. Sci. 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GAL4 binding domain and the herpes simplex virus transcriptional activation domain, VP16, and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. Proc. Natl. Acad. Sci. 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. J. Chem. Biol. 3:731-738; Fan, L. et al. 1999. Hum. Gene Ther. 10:2273-2285; Shariat, S.F. et al. 2001. Cancer Res. 61:2562-2571; Spencer, D.M. 1996. Curr. Biol. 6:839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known to those of skill in the art.

Other examples of cell or tissue-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A.M. et al., 2002. Circ Res. 91(12):1151-9); ubiquitin C promoter (Biochim Biophys Acta, 2003. Jan. 3;1625(1):52-63); Hsf2 promoter; murine COMP (cartilage oligomeric matrix protein) promoter; early B cell-specific mb-1 promoter (Sigvardsson M., et al., 2002. Mol. Cell Biol. 22(24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, J. Urol. 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. Proc. Natl. Acad. Sci. 99(24):15456-61); neural and liver ceramidase gene promoters (Okino N. et al., 2002. Biochem. Biophys. Res. Commun. 299(1):160-6); PSP94 gene promoter/enhancer (Gabril M.Y. et al., 2002. Gene Ther. 9(23):1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. Biol. Pharm. Bull. 25(11):1476-8); VL30 promoter (Staplin W.R. et al., 2002. Blood October 24, 2002); and IL-10 promoter (Brenner S., et al., 2002. J. Biol. Chem. December 18, 2002).

Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such as vitellogenin, very low-density lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavin-binding protein, biotin-binding protein (Awade, 1996. Z. Lebensm. Unters. Forsch. 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin (ovomacroglobin) promoters.

Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha-hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like growth factor I (IGF-I) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothioneine promoter, albumin promoter, and insulin promoter.

Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including, but not limited to, β lactoglobulin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

Promoters associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of CR2 (complement receptor type 2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, a steroid hormone-binding domain of the ovalbumin promoter is moved from about -6.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural backbone, are all encompassed by the current invention.

Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the present invention may be entire promoter sequences or fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a chicken ovalbumin promoter (SEQ ID NO:70). The constitutive and inducible promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:5).

As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-linked to their native promoter and may be located at any distance from their operably-linked promoter. A promoter operably-linked to an enhancer is referred to herein as an "enhanced promoter." The enhancers contained within the transposon-based vectors are preferably enhancers found in birds, and more preferably, an ovalbumin enhancer, but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter. In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a Chicken Ovalbumin enhancer from GenBank accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:71.

Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of an enhanced promoter or promoter. An exemplary 5' untranslated region is provided

in SEQ ID NO:72. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:73.

In one embodiment of the present invention, the first promoter operably-linked to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a tissue-specific promoter. In the second embodiment, use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is incorporated into the germline generally, the gene of interest is only expressed in a tissue-specific manner. A transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered by any route, and in one embodiment, the vector is administered to an ovary or to an artery leading to the ovary. In another embodiment, the vector is administered into the lumen of the oviduct or into an artery supplying the oviduct.

It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers to a majority of the expression of a particular gene of interest in the preferred cell or tissue, respectively.

When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter such as the ovalbumin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In a preferred embodiment, both the first promoter and the second promoter are an ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest or to an artery leading to the tissue of interest. In a preferred embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the lumen of the oviduct or an artery leading to the oviduct. In a

further preferred embodiment, administration is achieved by direct injection into the lumen of the magnum or the infundibulum of the oviduct.

Accordingly, cell specific promoters may be used to enhance transcription in selected tissues. In birds, for example, promoters that are found in cells of the 5 fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, 10 insulin, beta lactoglobin, whey acidic protein, lactalbumin, casein, and/or placental lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein production. Proteins made in the liver of birds may be delivered to the egg yolk.

15 In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For example, an avian actin promoter and its associated polyA sequence can be operably-linked to a transposase in a transposase-based vector for transfection into an avian. 20 Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

Directing Sequences

In some embodiments of the present invention, the gene of interest is 25 operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences. An egg directing sequence includes, but is not limited to, an ovomucoid signal sequence, an ovalbumin signal sequence, a cecropin prepro sequence, and a 30 vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid sequence, a portion or the entirety of which directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores

in the endoplasmic reticulum, or the plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the milk, when the transposon-based vectors are administered to a milk-producing animal.

5 Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. The present invention also includes a gene of interest operably-linked to a second gene 10 containing a signal sequence. An example of such an embodiment is wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence 15 including a sequence shown in SEQ ID NO:74. In another embodiment, the signal sequence is a shortened ovalbumin signal sequence including a sequence shown in SEQ ID NO:75 or SEQ ID NO:76.

As also used herein, the term "targeting sequence" refers to an amino acid 20 sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an 25 oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:77. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, 30 its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

Genes of Interest

The genes of interest in the polynucleotide cassette can be any gene, and preferably are genes that encode portions of multimeric proteins. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids

5 of the gene of interest, wherein the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the genes of interest are antibody genes or portions of antibody genes. Figure 2 shows a schematic drawing of a polynucleotide cassette containing an antibody heavy chain and an antibody light chain as two genes of interest. Antibodies used in or encoded by the

10 polynucleotide cassettes of the present invention include, but are not limited to, IgG, IgM, IgA, IgD, IgE, IgY, lambda chains, kappa chains, bi-specific antibodies, and fragments thereof; scFv fragments, Fc fragments, and Fab fragments as well as dimeric, trimeric and oligomeric forms of antibody fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies, human antibodies,

15 humanized antibodies, autoantibodies and hybrid antibodies. Genes encoding modified versions of naturally occurring antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies

20 against proteins associated with cancer-related molecules, such as anti-her 2, or anti-CA125. Accordingly, the present invention encompasses a polynucleotide cassette as described herein containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain.

Antibodies that may be produced using the present invention include, but are

25 not limited to, antibodies for use in cancer immunotherapy against specific antigens, or for providing passive immunity to an animal or a human against an infectious disease or a toxic agent. The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art. For example, antibodies may

30 be labeled with a fluorescent label attached that may be detected following exposure to specific wavelengths. Such labeled antibodies may be primary antibodies directed to a specific antigen, for example, rhodamine-labeled rabbit anti-growth hormone, or may be labeled secondary antibodies, such as fluorescein-labeled goat-anti chicken

IgG. Such labeled antibodies are known to one of ordinary skill in the art. The antibodies may also be designed to possess specific sequences useful for purification through means known to one of ordinary skill in the art. Labels useful for attachment to antibodies are also known to one of ordinary skill in the art. Some of these labels 5 are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Antibodies produced with the present invention may be used as laboratory reagents for numerous applications including radioimmunoassay, western blots, dot blots, ELISA, immunoaffinity columns and 10 other procedures requiring antibodies as known to one of ordinary skill in the art. Such antibodies include primary antibodies, secondary antibodies and tertiary antibodies, which may be labeled or unlabeled.

Additional antibodies that may be made with the practice of the present invention include, but are not limited to, primary antibodies, secondary antibodies, 15 designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, anti-hormone antibodies, anti-hypophysiotropic peptides, antibodies against non-natural antigens, anti-anterior pituitary hormone antibodies, anti-posterior pituitary hormone antibodies, anti-venom antibodies, anti-tumor marker antibodies, antibodies directed against epitopes associated with 20 infectious disease, including, anti-viral, anti-bacterial, antiprotozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotype, and anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme activity.

25 Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine γ -Globulin, Serum; Bovine IgG, Plasma; Chicken γ -Globulin, Serum; Human γ -Globulin, Serum; Human IgA, Plasma; Human IgA₁, Myeloma; Human IgA₂, Myeloma; Human IgA₂, Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human 30 IgG, Fab Fragment, Plasma; Human IgG, F(ab')₂ Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG₁, Myeloma; Human IgG₂, Myeloma; Human IgG₃, Myeloma; Human IgG₄, Myeloma; Human IgM, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chain κ , Urine; Human Immunoglobulin, Light

Chains κ and λ , Plasma; Mouse γ -Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit γ -Globulin, Serum; Rabbit IgG, Plasma; and Rat γ -Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding sequence of light and heavy chains of a murine monoclonal antibody that shows 5 specificity for human seminoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

A further non-limiting list of antibodies that recognize other antibodies and that may be produced using the present invention is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat γ -Globulin (Donkey); Anti-10 Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig γ -Globulin (Goat); Anti-Human Ig, Light Chain, Type κ Specific; Anti-Human Ig, Light Chain, Type λ Specific; Anti-Human IgA, α -Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE, ϵ -Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-15 Human IgG, Fc Fragment Specific (Goat); Anti-Human IgG, γ -Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG₁, Fc Fragment Specific; Anti-Human IgG₂, Fc Fragment Specific; Anti-Human IgG₂, Fd Fragment Specific; Anti-Human IgG₃, Hinge Specific; Anti-Human IgG₄, Fc Fragment Specific; 20 Anti-Human IgM, Fc Fragment Specific; Anti-Human IgM, μ -Chain Specific; Anti-Mouse IgE, ϵ -Chain Specific; Anti-Mouse γ -Globulin (Goat); Anti-Mouse IgG, γ -Chain Specific (Goat); Anti-Mouse IgG, γ -Chain Specific (Goat) F(ab')₂ Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM, μ -Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit γ -Globulin 25 (Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit IgG, H & L Chain Specific (Goat); Anti-Rat γ -Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey γ -Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

Antibodies that bind a particular ligand may also be produced. Exemplary ligands are as follows: adrenomedulin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide, interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neuropeptide Y, melanin concentrating hormone, neurokinin, substance P, leptin,

melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiostatin, bombesin, 5 endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin, osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, 10 vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokinetican, thyroid stimulating hormone, luteinizing hormone, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikriens, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:78 amino acid sequence, SEQ ID NO:79, nucleotide sequence), estrogen, testosterone, 15 corticosteroids, mineralocorticoids, thyroid hormone, thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, and prohormones, propeptides, splice variants, fragments and analogs thereof.

The following is yet another non-limiting of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab anti-platelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD18 antibody, anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, anti-hepatitis B antibody, anti-HER-2 antibody, anti-idiotype antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, anti-Rh 20 antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and Fab fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNF α , antibody directed against GPIIb/IIIa receptor on human platelets, gamma globulin, 25 anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against *S. aureus*, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the α subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the 30

IL-2 R α -chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recombinant antibody (IgG1k) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte 5 immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX-33 antibody to Fc γ R1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against 10 CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3) antibody, OKT3 ("muromomab-CD3") antibody, PRO 542 antibody, ReoPro ("abciximab") antibody, and TNF-IgG fusion protein.

Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix 15 Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs San Carlos CA, SIGMA, St.Louis, MO www.sigma-aldrich.com, Cappel ICN, Irvine, California, www.icnbiomed.com, and Calbiochem, La Jolla, California, www.calbiochem.com, which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies 20 may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the antibody polynucleotide sequence by choosing the codons that encode for each amino acid in the desired antibody.

Genes encoding protein and peptide hormones are a preferred class of genes of 25 interest in the present invention. Such protein and peptide hormones are synthesized throughout the endocrine system and include, but are not limited to, hypothalamic hormones and hypophysiotropic hormones, anterior, intermediate and posterior pituitary hormones, pancreatic islet hormones, hormones made in the gastrointestinal system, renal hormones, thymic hormones, parathyroid hormones, adrenal cortical 30 and medullary hormones. Specifically, hormones that can be produced using the present invention include, but are not limited to, chorionic gonadotropin, corticotropin, erythropoietin, glucagons, IGF-1, oxytocin, platelet-derived growth factor, calcitonin, follicle-stimulating hormone, luteinizing hormone, thyroid-

stimulating hormone, insulin, gonadotropin-releasing hormone and its analogs, vasopressin, octreotide, somatostatin, prolactin, adrenocorticotropic hormone, antidiuretic hormone, thyrotropin-releasing hormone (TRH), growth hormone-releasing hormone (GHRH), dopamine, melatonin, thyroxin (T₄), parathyroid 5 hormone (PTH), glucocorticoids such as cortisol, mineralocorticoids such as aldosterone, androgens such as testosterone, adrenaline (epinephrine), noradrenaline (norepinephrine), estrogens such as estradiol, progesterone, glucagons, calcitrol, calciferol, atrial-natriuretic peptide, gastrin, secretin, cholecystokinin (CCK), neuropeptide Y, ghrelin, PYY₃₋₃₆, angiotensinogen, thrombopoietin, and leptin. By 10 using appropriate polynucleotide sequences, species-specific hormones may be made by transgenic animals.

In one embodiment of the present invention, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become 15 linked together to form the insulin molecule. In these embodiments, proinsulin is expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:80, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65.

Further included in the present invention are genes of interest that encode 20 proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues (GALT) system. The immune system proteins and peptides that can be made in transgenic animals using the polynucleotide cassettes of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon, 25 alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interlukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF- α , and TNF- β . Other cytokines included in the present invention include cardiotrophin, stromal cell derived factor, macrophage derived chemokine (MDC), melanoma growth 30 stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1 alpha), 2, 3 alpha, 3 beta, 4 and 5.

Genes encoding lytic peptides such as p146 are also included in the genes of interest of the present invention. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:78. The present invention also encompasses a

polynucleotide cassette comprising a p146 nucleic acid having a sequence of SEQ ID NO:79.

Enzymes are another class of proteins that may be encoded by the polynucleotide cassettes of the present invention. Such enzymes include but are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnaseI, hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase and plasminogen activator. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in activation of the enzyme.

Extracellular matrix proteins are one class of desired proteins that may be encoded by the polynucleotide cassettes of the present invention. Examples include but are not limited to collagen, fibrin, elastin, laminin, and fibronectin and subtypes thereof. Intracellular proteins and structural proteins are other classes of desired proteins in the present invention.

Growth factors are another desired class of proteins that may be encoded by the polynucleotide cassettes of the present invention and include, but are not limited to, transforming growth factor- α ("TGF- α "), transforming growth factor- β (TGF- β), platelet-derived growth factors (PDGF), fibroblast growth factors (FGF), including FGF acidic isoforms 1 and 2, FGF basic form 2 and FGF 4, 8, 9 and 10, nerve growth factors (NGF) including NGF 2.5s, NGF 7.0s and beta NGF and neurotrophins, brain derived neurotrophic factor, cartilage derived factor, growth factors for stimulation of the production of red blood cells, growth factors for stimulation of the production of white blood cells, bone growth factors (BGF), basic fibroblast growth factor, vascular endothelial growth factor (VEGF), granulocyte colony stimulating factor (G-CSF), insulin like growth factor (IGF) I and II, hepatocyte growth factor, glial neurotrophic growth factor (GDNF), stem cell factor (SCF), keratinocyte growth factor (KGF), transforming growth factors (TGF), including TGFs alpha, beta, beta1, beta2, beta3, skeletal growth factor, bone matrix derived growth factors, bone derived growth factors, erythropoietin (EPO) and mixtures thereof.

Another desired class of proteins that may be encoded by the polynucleotide cassettes of the present invention include, but are not limited to, leptin, leukemia inhibitory factor (LIF), tumor necrosis factor alpha and beta, ENBREL, angiostatin, endostatin, thrombospondin, osteogenic protein-1, bone morphogenetic proteins 2 and 5 7, osteonectin, somatomedin-like peptide, and osteocalcin.

Yet another desired class of proteins encoded by the genes of interest are blood proteins or clotting cascade protein including albumin, Prekallikrein, High molecular weight kininogen (HMWK) (contact activation cofactor; Fitzgerald, Flaujeac Williams factor), Factor I (Fibrinogen), Factor II (prothrombin), Factor III (Tissue 10 Factor), Factor IV (calcium), Factor V (proaccelerin, labile factor, accelerator (Ac-) globulin), Factor VI (Va) (accelerin), Factor VII (proconvertin), serum prothrombin conversion accelerator (SPCA), cothromboplastin), Factor VIII (antihemophilic factor A, antihemophilic globulin (AHG)), Factor IX (Christmas Factor, antihemophilic factor B, plasma thromboplastin component (PTC)), Factor X (Stuart- 15 Prower Factor), Factor XI (Plasma thromboplastin antecedent (PTA)), Factor XII (Hageman Factor), Factor XIII (rotransglutaminase, fibrin stabilizing factor (FSF), fibrinoligase), von Willebrand factor, Protein C, Protein S, Thrombomodulin, Antithrombin III.

A non-limiting list of the peptides and proteins that may be encoded by the 20 polynucleotide cassettes of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA, (St.Louis, MO www.sigma-aldrich.com), Cappel ICN (Irvine, California, www.icnbiomed.com), and Calbiochem (La Jolla, California, www.calbiochem.com). 25 The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired protein or peptide.

30 Some of these desired proteins or peptides that may be encoded by the polynucleotide cassettes of the present invention include but are not limited to the following: adrenomedulin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide,

interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone, melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, leumorphin, 5 peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiotatin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony 10 stimulating factor (GM-CSF), motilin, neuropeptid Y, proopiomelanocortin, osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokinetican, thyroid stimulating hormone, luteinizing hormone, follicle stimulating 15 hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikreins, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:78, amino acid sequence, SEQ ID NO:79, nucleotide sequence), thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, apolipoproteins, apolipoprotein A, apolipoprotein B, and prohormones, propeptides, splice variants, fragments and 20 analogs thereof.

Other desired proteins that may be encoded by the polynucleotide cassettes of the present invention include bacitracin, polymixin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, anti-hepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) 25 antibody, adenosine deaminase, anti-digoxin antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodetus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin, α -galactosidase, gramicidin, 30 streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3 ("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide,

lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on platelets (integrilin), goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinin, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of 5 influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor 10 (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP), humanized mAb to rous sarcoma virus (MEDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsonin enzyme, tissue 15 factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNF α , mAb against factor VII, relaxin, capreomycin, glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, alteplase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectaplate, anti-CD11a antibody (hu 1124), 20 anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, Fab fragment of anti-VEGF Ab, APO2 ligand (tumor necrosis factor-related apoptosis-inducing ligand), rTGF- β (transforming growth factor- β), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase, α -L-iduronidase, α -L-galactosidase (humanglutamic acid decarboxylase, 25 acid sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (IR501), BI 1050/1272 mAb against very late antigen-4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotype antibody to GD3 glycolipid, antibacterial peptide against *H. pylori*, MDX-447 bispecific humanized mAb to EGF 30 receptor, anti-cytomegalovirus (CMV), Medi-491 B19 parvovirus vaccine, humanized recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on *Escherechia coli* strains), proteins of lyme disease vaccine against *B. burgdorferi* protein (DbpA), proteins of Medi-501 human

papilloma virus-11 vaccine (HPV), *Streptococcus pneumoniae* vaccine, Medi-507 mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb to Fc γ R1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bi-specific mAb against HER-2, MDX-447 bi-specific mAb against EGF receptor, 5 MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R α -chain (basiliximab), mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A), humanized mAb against the epidermal growth receptor-2, humanized mAb against the α subunit of the interleukin-2 receptor, low molecular weight heparin, anti-10 hemophilic factor, and bactericidal/permeability-increasing protein (r-BPI).

Other multimeric proteins that may be produced using the present invention are as follows: factors involved in the synthesis or replication of DNA, such as DNA polymerase alpha and DNA polymerase delta; proteins involved in the production of mRNA, such as TFIID and TFIIH; cell, nuclear and other membrane-associated 15 proteins, such as hormone and other signal transduction receptors, active transport proteins and ion channels, multimeric proteins in the blood, including hemoglobin, fibrinogen and von Willabrand's Factor; proteins that form structures within the cell, such as actin, myosin, and tubulin and other cytoskeletal proteins; proteins that form structures in the extra cellular environment, such as collagen, elastin and fibronectin; 20 proteins involved in intra- and extra-cellular transport, such as kinesin and dynein, the SNARE family of proteins (soluble NSF attachment protein receptor) and clathrin; proteins that help regulate chromatin structure, such as histones and protamines, Swi3p, Rsc8p and moira; multimeric transcription factors such as Fos, Jun and CBTF (CCAAT box transcription factor); multimeric enzymes such as acetylcholinesterase 25 and alcohol dehydrogenase; chaperone proteins such as GroE, Gro EL (chaperonin 60) and Gro ES(chaperonin 10); anti-toxins, such as snake venom, botulism toxin, *Streptococcus* super antigens; lysins (enzymes from bacteriophage and viruses); as well as most allosteric proteins.

The multimeric proteins made using the present invention may be labeled 30 using labels and techniques known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Some of these labels may be genetically

engineered into the polynucleotide sequence for the expression of the selected multimeric protein. The peptides and proteins may also have label-incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label multimeric protein.

5 It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

10 The present invention may also be used to produce desired molecules other than proteins and peptides including, but not limited to, lipoproteins such as high density lipoprotein (HDL), HDL-Milano, and low density lipoprotein, lipids, carbohydrates, siRNA and ribozymes. In these embodiments, a gene of interest encodes a nucleic acid molecule or a protein that directs production of the desired molecule.

15 The present invention further encompasses the use of inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. These inhibitory molecules include antisense nucleic acids, siRNA and inhibitory proteins. In a preferred embodiment, the endogenous protein whose expression is inhibited is an egg white protein including, but not limited to ovalbumin, ovotransferrin, and ovomucin
20 ovomucoid, ovoinhibitor, cystatin, ovostatin, lysozyme, ovoglobulin G2, ovoglobulin G3, avidin, and thiamin binding protein. In one embodiment, a polynucleotide cassette containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, is transfected into an animal such as a bird and the bird's production of endogenous ovalbumin protein is reduced by the interference
25 RNA mechanism (RNAi). In other embodiments, a polynucleotide cassette encodes an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production. Endogenous egg white production can be inhibited in an avian at any time, but is
30 preferably inhibited preceding, or immediately preceding, the harvest of eggs.

Modified Desired Proteins and Peptides

The present invention may be used for the production of multimeric proteins. "Proteins", "peptides," "polypeptides" and "oligopeptides" are chains of amino acids

(typically L-amino acids) whose alpha carbons are linked through peptide bonds formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein.

5 Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the carboxyl group of an amino acid at any other location within the protein.

10

15

Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may 20 encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide bond mimetic includes peptide backbone modifications well known to those skilled in the art.

Furthermore, one of skill will recognize that, as mentioned above, individual 25 substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino 30 acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);

- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

5 A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid being substituted. Thus, the substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A “conservative substitution” also refers to utilizing a substituting amino acid which is identical to the amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

10 Suitable protecting groups are described in Green and Wuts, “Protecting Groups in Organic Synthesis”, John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. J. Pharm. Sci. 57:783; Ditter et al., 1968. J. Pharm. Sci. 57:828; Ditter et al., 1969. J. Pharm. Sci. 58:557; King et al., 1987. Biochemistry 26:2294; Lindberg et al., 1989. Drug Metabolism and Disposition 17:311; Tunek et al., 1988. Biochem. Pharm. 37:3867; Anderson et al., 1985 Arch. Biochem. Biophys. 239:538; and Singhal et al., 1987. FASEB J. 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups 15 and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, 20 ethyl, benzyl or substituted benzyl ester, more preferably as a benzyl ester.

25 Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another

amino acid from the same group. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

Group I includes leucine, isoleucine, valine, methionine and modified amino acids
5 having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine.

Group II includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine.

Group III includes phenylalanine, phenylglycine, tyrosine, tryptophan, 10 cyclohexylmethyl glycine, and modified amino residues having substituted benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro, —NH₂, methoxy, ethoxy and —CN. Preferably, Group III includes phenylalanine, tyrosine and tryptophan.

Group IV includes glutamic acid, aspartic acid, a substituted or unsubstituted 15 aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine, —CO—NH— alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the side chain —(CH₂)₃—COOH, an ester thereof (substituted or unsubstituted 20 aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and asparagine.

25 Group V includes histidine, lysine, ornithine, arginine, N-nitroarginine, β -cycloarginine, γ -hydroxyarginine, N-amidinocitruline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine, arginine and ornithine. A homolog of an amino acid includes from 1 to about 30 3 additional or subtracted methylene units in the side chain.

Group VI includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side chains substituted with —OH or —SH, for

example, —CH₂CH₂OH, —CH₂CH₂CH₂OH or -CH₂CH₂OHCH₃. Preferably, Group VI includes serine, cysteine or threonine.

In another aspect, suitable substitutions for amino acid residues include “severe” substitutions. A “severe substitution” is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for glycine, a D amino acid for the corresponding L amino acid, or —NH—CH[(-(CH₂)₅—COOH)—CO— for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties that the functional group of the amino acid being substituted. Examples of such modifications include tryptophan for glycine, lysine for aspartic acid and —(CH₂)₄COOH for the side chain of serine. These examples are not meant to be limiting.

In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the following manner:

AA₁ is serine, glycine, alanine, cysteine or threonine;
AA₂ is alanine, threonine, glycine, cysteine or serine;
AA₃ is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β -cycloarginine, γ -hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
AA₄ is proline, leucine, valine, isoleucine or methionine;
AA₅ is tryptophan, alanine, phenylalanine, tyrosine or glycine;

AA₆ is serine, glycine, alanine, cysteine or threonine;
AA₇ is proline, leucine, valine, isoleucine or methionine;
AA₈ is alanine, threonine, glycine, cysteine or serine;
AA₉ is alanine, threonine, glycine, cysteine or serine;

5 AA₁₀ is leucine, isoleucine, methionine or valine;
AA₁₁ is serine, glycine, alanine, cysteine or threonine;
AA₁₂ is leucine, isoleucine, methionine or valine;
AA₁₃ is leucine, isoleucine, methionine or valine;
AA₁₄ is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or
10 unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
AA₁₅ is arginine, N-nitroarginine, β -cycloarginine, γ -hydroxy-arginine, N-
amidinocitruline or 2-amino-4-guanidino-butanoic acid
AA₁₆ is proline, leucine, valine, isoleucine or methionine;
AA₁₇ is serine, glycine, alanine, cysteine or threonine;

15 AA₁₈ is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or
unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
AA₁₉ is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine,
methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or
aspartic acid;

20 AA₂₀ is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-
nitroarginine, β -cycloarginine, γ -hydroxyarginine, N-amidinocitruline or 2-amino-4-
guanidinobutanoic acid;
AA₂₁ is alanine, threonine, glycine, cysteine or serine;
AA₂₂ is alanine, threonine, glycine, cysteine or serine;

25 AA₂₃ is histidine, serine, threonine, cysteine, lysine or ornithine;
AA₂₄ is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted
aliphatic or aryl ester of glutamic acid or aspartic acid;
AA₂₅ is asparagine, aspartic acid, glutamic acid, glutamine, leucine, valine,
isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of
30 glutamic acid or aspartic acid; and
AA₂₆ is cysteine, histidine, serine, threonine, lysine or ornithine.

It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

In one embodiment of the present invention, codons for the first several N-terminal amino acids of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal codons of the gene of interest are modified in this manner.

When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, as with the multivalent multimeric proteins, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also be contained within a nucleotide sequence with a purification handle or be flanked by proteolytic cleavage sites.

Such polypeptide spacers may have from about 1 to about 100 amino acids, preferably 3 to 20 amino acids, and more preferably 4-15 amino acids. The spacers in a polypeptide are independently chosen, but are preferably all the same. The spacers should allow for flexibility of movement in space and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)_x (SEQ ID NO:81) and (Gly₄-Ser)_y, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

(Gly-Pro-Gly-Gly)₃

SEQ ID NO:82 Gly Pro Gly Gly Gly Pro Gly Gly Gly Pro Gly Gly

(Gly₄-Ser)₃

SEQ ID NO:83 Gly Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser

5 or (Gly₄-Ser)₄

SEQ ID NO:84 Gly Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
Gly Gly Gly Gly Ser.

One example of a multivalent multimeric protein containin a spacer is leutinizing hormone (LH), normally made as separate alpha and beta chains, made as 10 a single polypeptide as described in Galet et. al., Mol. Cell Endocrinology, 2001, 174 (1-2):31-40. Production of a multimeric protein may thus be simplified using a spacer sequence that may or may not contain cleavage sites. In the case of an immunoglobulin, for example, a heavy and light chain may be synthesized as a single polypeptide using a spacer sequence with protease sites native to the transgenic 15 animal so as to make, upon processing, a heavy and light chain combination in close association, facilitating the addition of a similar heavy and light chain to produce the native immunoglobulin. In this model, the removal of the spacer sequence may or may not be required. Other multimeric proteins may be made in bioengineered organisms in a similar fashion.

20 Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein may also be built into the vector. Such sequences are known in the art and include the glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and maltose binding protein.

25 Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal, specific nucleotide sequences encoding for amino acids recognized by enzymes may 30 be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest.

Nucleotide sequences encoding other amino acid sequences which display pH sensitivity, chemical sensitivity or photolability may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thrombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the definition of cleavage site as used herein. Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid. Self-splicing cleavage sites such as inteins are also included in the present invention.

In some embodiments, one or more cleavage sites are incorporated into a polynucleotide cassette containing multiple genes of interest. Figure 4 depicts one example of a polynucleotide cassette containing two genes of interest containing a cleavage site between them. The genes of interest may encode different proteins or peptides, the same protein or peptide, or modified versions of the same protein or peptide. While Figure 4 shows a polynucleotide cassette containing two genes of interest, the present invention encompasses a polynucleotide cassette containing any number of genes of interest. The cleavage site located between the genes of interest can encode any amino acid sequence that is cleaved by any means. As mentioned above, the cleavage site can encode an amino acid sequence cleaved by a protease, a chemical reaction, can be a photolabile site, can be a pro polynucleotide

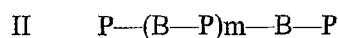
The present invention includes a polynucleotide cassette that encodes a repetitive polypeptide chain in which two or more peptides, polypeptides or proteins, designated as P in the structural formulae presented below, are each separated by a peptide spacer or cleavage site designated as B. A polypeptide multivalent ligand, also called a multivalent protein, is a form of a multimeric protein encoded by the polynucleotide cassettes of the present invention, and is represented by structural formulae (I, II and III). Each peptide or protein is connected to another peptide or protein through a peptide bond, to a linker group, to a spacer, or to a cleavage site. Each peptide, polypeptide or protein may be the same or different and each linker, spacer, cleavage site or covalent bond is independently chosen.

A "polypeptide multivalent protein" is a multiple repeat polypeptide chain in which two or more peptides P are each separated by a peptide linker group, a spacer or a cleavage site. A polypeptide multivalent ligand is represented by structural formulae II and III.



wherein B is a peptide spacer or cleavage site, n is an integer from 2 to about 20, each L is a covalent bond, a linking group or cleavage site which may be present or absent, and each P is a peptide having from about 4 to about 200 amino acid residues.

10



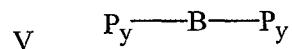
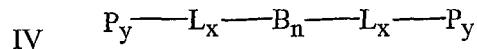
wherein m is an integer from 0 to about 20.



wherein n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7, further wherein a is 1.

Other examples of multivalent proteins include the following:

20



In the preceding structural formulae IV and V of polypeptide multivalent ligands encoded by a polynucleotide cassette of the present invention, each P is a peptide having from about 4 to about 200 amino acid residues, y is 1, x is an integer from 1 to 3, and n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7. Each B 5 is a peptide spacer or cleavage site comprised of at least 2 amino acids or a cleavage site. Each peptide P and each B are independently chosen and may be the same or different.

Suitable linkers (L) are groups that can connect peptides and proteins to each other. In one example, the linker is an oligopeptide of from about 1 to about 10 10 amino acids consisting of amino acids with inert side chains. Suitable oligopeptides include polyglycine, polyserine, polyproline, polyalanine and oligopeptides consisting of alanyl and/or serinyl and/or prolinyl and/or glycyl amino acid residues. m in structural formula II is an integer from 0 to about 20.

The peptides, polypeptides and proteins in a multivalent protein can be 15 connected to each other by covalent bonds, linker groups, spacers, cleavage groups or a combination thereof. The linking groups can be the same or different.

A polypeptide spacer shown in structural formula (II) is a peptide having from about 5 to about 40 amino acid residues. The spacers in a polypeptide multivalent ligand are independently chosen, and may be the same or different. The spacers 20 should allow for flexibility of movement in space for the flanking peptides. polypeptides and proteins P, and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers 25 are (Gly-Pro-Gly-Gly)_x (SEQ ID NO:81) and (Gly₄-Ser)_y, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include (Gly₄-Ser)₃ (SEQ ID NO:82). Spacers can also include from one to about four amino acids that create a signal for proteolytic cleavage.

30 In another embodiment of the present invention, a TAG sequence is linked to a gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to

isolate the protein using affinity purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly, as used herein, a TAG sequence includes a spacer sequence, a purification handle and a cleavage site. The spacer sequences in the TAG proteins contain one or more 5 repeats shown in SEQ ID NO:85. A preferred spacer sequence comprises the sequence provided in SEQ ID NO:86. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide sequences are provided in SEQ ID NO:87 and SEQ ID NO:88, respectively. However, it should be understood that any antigenic region, or otherwise associative 10 regions such as avidin/biotin, may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ ID NO:89) and a furin cleavage 15 site. In one embodiment of the present invention, the TAG sequence comprises a polynucleotide sequence of SEQ ID NO:90.

Methods of Administering Polynucleotide Cassettes

In addition to the polynucleotide cassettes described above, the present invention also includes methods of administering the polynucleotide cassettes to an 20 animal, methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic animal wherein a gene of interest is incorporated into cells other than the germline cells of the animal. The polynucleotide cassettes may reside in any vector or delivery solution when administered or may be naked DNA. In one embodiment, a 25 transposon-based vector containing the polynucleotide cassette between two insertion sequences recognized by a transposase is administered to an animal. The polynucleotide cassettes of the present invention may be administered to an animal via any method known to those of skill in the art, including, but not limited to, intraembryonic, intratesticular, intraoviduct, intraovarian, into the duct system of the 30 mammary gland, intraperitoneal, intraarterial, intravenous, topical, oral, nasal, and pronuclear injection methods of administration, or any combination thereof. The polynucleotide cassettes may also be administered within the lumen of an organ, into an organ, into a body cavity, into the cerebrospinal fluid, through the urinary system,

through the genitourinary system, through the reproductive system, or through any route to reach the desired cells.

The polynucleotide cassettes may be delivered through the vascular system to be distributed to the cells supplied by that vessel. For example, the compositions may 5 be placed in the artery supplying the ovary or supplying the fallopian tube to transfect cells in those tissues. In this manner, follicles could be transfected to create a germline transgenic animal. Alternatively, supplying the compositions through the artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells could manufacture a desired protein or peptide 10 for deposition in the egg white. Administration of the compositions through the portal vein would target uptake and transformation of hepatic cells. Administration through the urethra and into the bladder would target the transitional epithelium of the bladder. Administration through the vagina and cervix would target the lining of the uterus and the epithelial cells of the fallopian tube. Administration through the internal 15 mammary artery or through the duct system of the mammary gland would transfect secretory cells of the lactating mammary gland to perform a desired function, such as to synthesize and secrete a desired protein or peptide into the milk.

The polynucleotide cassettes may be administered in a single administration, multiple administrations, continuously, or intermittently. The polynucleotide 20 cassettes may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, a polynucleotide cassette is administered to an animal in multiple administrations, each administration containing the polynucleotide cassette and a different transfecting reagent.

In a preferred embodiment, the animal is an egg-laying animal, and more 25 preferably, an avian. In one embodiment, between approximately 1 and 150 µg, 1 and 100 µg, 1 and 50 µg, preferably between 1 and 20 µg, and more preferably between 5 and 10 µg of a transposon-based vector containing the polynucleotide cassette is administered to the oviduct of a bird. In a chicken, it is preferred that between approximately 1 and 100 µg, or 5 and 50 µg are administered. In a quail, it is 30 preferred that between approximately 5 and 10 µg are administered. Optimal ranges depending upon the type of bird and the bird's stage of sexual maturity. Intraoviduct administration of the transposon-based vectors of the present invention result in a PCR positive signal in the oviduct tissue, whereas intravascular administration results

in a PCR positive signal in the liver. In other embodiments, the polynucleotide cassettes is administered to an artery that supplies the oviduct or the liver. These methods of administration may also be combined with any methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of 5 naked DNA, and use of dimethyl sulfoxide (DMSO).

The transposon-based vectors may be administered to the animal at any point during the lifetime of the animal, however, it is preferable that the vectors are administered prior to the animal reaching sexual maturity. The transposon-based vectors are preferably administered to a chicken oviduct between approximately 14 10 and 16 weeks of age and to a quail oviduct between approximately 5 and 10 weeks of age, more preferably 5 and 8 weeks of age, and most preferably between 5 and 6 weeks of age, when standard poultry rearing practices are used. The vectors may be administered at earlier ages when exogenous hormones are used to induce early sexual maturation in the bird. In some embodiments, the transposon-based vector is 15 administered to an animal's oviduct following an increase in proliferation of the oviduct epithelial cells and/or the tubular gland cells. Such an increase in proliferation normally follows an influx of reproductive hormones in the area of the oviduct. When the animal is an avian, the transposon-based vector is administered to the avian's oviduct following an increase in proliferation of the oviduct epithelial cells 20 and before the avian begins to produce egg white constituents.

The present invention also includes a method of intraembryonic administration of a transposon-based vector containing a polynucleotide cassette to an avian embryo comprising the following steps: 1) incubating an egg on its side at room temperature for two hours to allow the embryo contained therein to move to top dead center 25 (TDC); 2) drilling a hole through the shell without penetrating the underlying shell membrane; 3) injecting the embryo with the transposon-based vector in solution; 4) sealing the hole in the egg; and 5) placing the egg in an incubator for hatching. Administration of the transposon-based vector can occur anytime between immediately after egg lay (when the embryo is at Stage X) and hatching. Preferably, 30 the transposon-based vector is administered between 1 and 7 days after egg lay, more preferably between 1 and 2 days after egg lay. The transposon-based vectors may be introduced into the embryo in amounts ranging from about 5.0 μ g to 10 pg, preferably 1.0 μ g to 100 pg. Additionally, the transposon-based vector solution volume may be

between approximately 1 μ l to 75 μ l in quail and between approximately 1 μ l to 500 μ l in chicken.

The present invention also includes a method of intratesticular administration of a transposon-based vector containing a polynucleotide cassette including injecting 5 a bird with a composition comprising the transposon-based vector, an appropriate carrier and an appropriate transfection reagent. In one embodiment, the bird is injected before sexual maturity, preferably between approximately 4-14 weeks, more preferably between approximately 6-14 weeks and most preferably between 8-12 weeks old. In another embodiment, a mature bird is injected with a transposon-based 10 vector an appropriate carrier and an appropriate transfection reagent. The mature bird may be any type of bird, but in one example the mature bird is a quail.

A bird is preferably injected prior to the development of the blood-testis barrier, which thereby facilitates entry of the transposon-based vector into the seminiferous tubules and transfection of the spermatogonia or other germline cells. 15 At and between the ages of 4, 6, 8, 10, 12, and 14 weeks, it is believed that the testes of chickens are likely to be most receptive to transfection. In this age range, the blood/testis barrier has not yet formed, and there is a relatively high number of spermatogonia relative to the numbers of other cell types, e.g., spermatids, etc. See J. Kumaran et al., 1949. Poultry Sci., 29:511-520. See also E. Oakberg, 1956. Am. J. 20 Anatomy, 99:507-515; and P. Kluin et al., 1984. Anat. Embryol., 169:73-78.

The transposon-based vectors may be introduced into a testis in an amount ranging from about 0.1 μ g to 10 μ g, preferably 1 μ g to 10 μ g, more preferably 3 μ g to 10 μ g. In a quail, about 5 μ g is a preferred amount. In a chicken, about 5 μ g to 10 μ g per testis is preferred. These amounts of vector DNA may be injected in one dose or 25 multiple doses and at one site or multiple sites in the testis. In a preferred embodiment, the vector DNA is administered at multiple sites in a single testis, both testes being injected in this manner. In one embodiment, injection is spread over three injection sites: one at each end of the testis, and one in the middle. Additionally, the transposon-based vector solution volume may be between approximately 1 μ l to 30 75 μ l in quail and between approximately 1 μ l to 500 μ l in chicken. In a preferred embodiment, the transposon-based vector solution volume may be between approximately 20 μ l to 60 μ l in quail and between approximately 50 μ l to 250 μ l in

chicken. Both the amount of vector DNA and the total volume injected into each testis may be determined based upon the age and size of the bird.

According to the present invention, the polynucleotide cassette is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed include, but are not limited to, the following: cationic lipid transfection reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof; SUPERFECT®, Cytofectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin, DMRIE-C oligofectamine, and PLUS reagent from InVitrogen; Xtreme gene, fugene, DOSPER and DOTAP from Roche; Lipotaxi and Genejammer from Strategene; and Escort from SIGMA. In one embodiment, the transfection reagent is SUPERFECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, a transposon-based vector containing a polynucleotide cassette is administered intratesticularly and the ratio of DNA to transfection reagent can be from 1:1.5 to 1:15, preferably 1:2 to 1:10, all expressed as wt/vol. Transfection may also be accomplished using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

Testing for and Breeding Animals Carrying the Transgene

Following administration of a polynucleotide cassette to an animal, DNA is extracted from the animal to confirm integration of the genes of interest. Advantages provided by the present invention include the high rates of integration, or 5 incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovary route (including intraarterial administrations to arteries leading to the oviduct or ovary) and contained within a transposon-based vector.

Actual frequencies of integration can be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by 10 quantitative PCR. Another method for estimating the rate of transgene insertion is the so-called primed in situ hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of 15 PCR, and the slides are then developed through normal in situ hybridization procedures. This technique combines the best features of in situ PCR and fluorescence in situ hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question. The 28s rRNA gene will be used as a positive control for spermatogonia to confirm that the technique is functioning properly. 20 Using different fluorescent labels for the transgene and the 28s gene causes cells containing a transgene to fluoresce with two different colored tags.

Breeding experiments may also be conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment performed according to the present invention, each male bird is exposed to 2-3 25 different adult female birds for 3-4 days each. This procedure is continued with different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

30 Any male producing a transgenic embryo is bred to additional females. Eggs from these females are incubated, hatched, and the chicks tested for the exogenous DNA. Any embryos that die are necropsied and examined directly for the transgene or protein encoded by the transgene, either by fluorescence or PCR. The offspring

that hatch and are found to be positive for the exogenous DNA are raised to maturity. These birds are bred to produce further generations of transgenic birds, to verify efficiency of the transgenic procedure and the stable incorporation of the transgene into the germ line. The resulting embryos are examined for transgene presence at day 5 3 or 4 using PCR.

It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be performed to amplify gene copy numbers and protein output.

Production of Desired Multimeric Proteins in Egg White

10 In one embodiment, a transposon-based vector containing a polynucleotide cassette of the present invention may be administered to a bird for production of desired proteins or peptides in the egg white. These transposon-based vectors preferably contain one or more of an ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific 15 ovalbumin promoters are described in B. O'Malley et al., 1987. EMBO J., vol. 6, pp. 2305-12; A. Qiu et al., 1994. Proc. Nat. Acad. Sci. (USA), vol. 91, pp. 4451-4455; D. Monroe et al., 2000. Biochim. Biophys. Acta, 1517 (1):27-32; H. Park et al., 2000. Biochem., 39:8537-8545; and T. Muramatsu et al., 1996. Poult. Avian Biol. Rev., 6:107-123.

20 Production of Desired Multimeric Proteins in Egg Yolk

The present invention is particularly advantageous for production of recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins, 25 lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant protein production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and re-folding of the recombinantly-produced proteins, which may deleteriously affect their 30 structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in sufficiently high amounts in recombinant protein production systems. The present

invention provides an advantageous resolution of the problem of protein and peptide solubility during production of large amounts of recombinant proteins.

In one embodiment of the present invention, deposition of a desired protein into the egg yolk is accomplished by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a gene of interest that encodes a desired protein. This polynucleotide cassette can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:77.

In a preferred embodiment, the transposon-based vector contains a transposase gene operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

25 Isolation and Purification of Desired Multimeric Proteins

For large-scale production of protein, an animal breeding stock that is homozygous for the transgene is preferred. Such homozygous individuals are obtained and identified through, for example, standard animal breeding procedures or PCR protocols.

30 Once expressed, peptides, polypeptides and proteins can be purified according to standard procedures known to one of ordinary skill in the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis, high performance liquid chromatography, immunoprecipitation and the like.

Substantially pure compositions of about 50 to 99% homogeneity are preferred, and 80 to 95% or greater homogeneity are most preferred for use as therapeutic agents.

In one embodiment of the present invention, the animal in which the desired protein is produced is an egg-laying animal. In a preferred embodiment of the present invention, the animal is an avian and a desired peptide, polypeptide or protein is isolated from an egg white. Egg white containing the exogenous protein or peptide is separated from the yolk and other egg constituents on an industrial scale by any of a variety of methods known in the egg industry. See, e.g., W. Stadelman et al. (Eds.), Egg Science & Technology, Haworth Press, Binghamton, NY (1995). Isolation of the exogenous peptide or protein from the other egg white constituents is accomplished by any of a number of polypeptide isolation and purification methods well known to one of ordinary skill in the art. These techniques include, for example, chromatographic methods such as gel permeation, ion exchange, affinity separation, metal chelation, HPLC, and the like, either alone or in combination. Another means that may be used for isolation or purification, either in lieu of or in addition to chromatographic separation methods, includes electrophoresis. Successful isolation and purification is confirmed by standard analytic techniques, including HPLC, mass spectroscopy, and spectrophotometry. These separation methods are often facilitated if the first step in the separation is the removal of the endogenous ovalbumin fraction of egg white, as doing so will reduce the total protein content to be further purified by about 50%.

To facilitate or enable purification of a desired protein or peptide, the polynucleotide cassettes may include one or more additional epitopes or domains. Such epitopes or domains include DNA sequences encoding enzymatic, chemical or photolabile cleavage sites including, but not limited to, an enterokinase cleavage site; the glutathione binding domain from glutathione S-transferase; polylysine; hexahistidine or other cationic amino acids, and sites cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid; thioredoxin; hemagglutinin antigen; maltose binding protein; a fragment of gp41 from HIV; and other purification epitopes or domains commonly known to one of skill in the art. Other proteolytic cleavage sites that may be included in the polynucleotide cassettes are cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-

protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thrombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Self-splicing cleavage sites such as inteins may also be included in the polynucleotide cassettes of the present invention.

5 In one representative embodiment, purification of desired proteins from egg white utilizes the antigenicity of the ovalbumin carrier protein and particular attributes of a TAG linker sequence that spans ovalbumin and the desired protein. The TAG sequence is particularly useful in this process because it contains 1) a highly antigenic epitope, a fragment of gp41 from HIV, allowing for stringent affinity purification,
10 and, 2) a recognition site for the protease enterokinase immediately juxtaposed to the desired protein. In a preferred embodiment, the TAG sequence comprises approximately 50 amino acids. A representative TAG sequence is provided below.

15 Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys
Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys (SEQ ID NO:90)

20 The underlined sequences were taken from the hairpin loop domain of HIV gp-41 (SEQ ID NO:87). Sequences in italics represent the cleavage site for enterokinase (SEQ ID NO:89). The spacer sequence upstream of the loop domain was made from repeats of (Pro Ala Asp Asp Ala) (SEQ ID NO:85) to provide free rotation and promote surface availability of the hairpin loop from the ovalbumin carrier protein.

Isolation and purification of a desired protein is performed as follows:

25 1. Enrichment of the egg white protein fraction containing ovalbumin and the transgenic ovalbumin-TAG-desired protein.

2. Size exclusion chromatography to isolate only those proteins within a narrow range of molecular weights (a further enrichment of step 1).

3. Ovalbumin affinity chromatography. Highly specific antibodies to ovalbumin will eliminate virtually all extraneous egg white proteins except ovalbumin
30 and the transgenic ovalbumin-TAG-desired protein.

4. gp41 affinity chromatography using anti-gp41 antibodies. Stringent application of this step will result in virtually pure transgenic ovalbumin-TAG-desired protein.

5. Cleavage of the transgene product can be accomplished in at least one of two ways:

- 5 a. The transgenic ovalbumin-TAG-desired protein is left attached to the gp41 affinity resin (beads) from step 4 and the protease enterokinase is added. This liberates the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence is retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), leaves the desired protein together with enterokinase in solution. Enterokinase is recovered and reused.
- 10 b. Alternatively, enterokinase is immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-desired protein eluted from the affinity column of step 4 is then applied to the protease resin. Protease action cleaves the ovalbumin-TAG sequence from the desired protein and leaves both entities in solution. The immobilized enterokinase resin is recharged and reused.
- 15 c. The choice of these alternatives is made depending upon the size and chemical composition of the transgene target protein.

6. A final separation of either of these two (5a or 5b) protein mixtures is made
20 using size exclusion, or enterokinase affinity chromatography. This step allows for desalting, buffer exchange and/or polishing, as needed.

Cleavage of the transgene product (ovalbumin-TAG-desired protein) by enterokinase, then, results in two products: ovalbumin-TAG and the desired protein. More specific methods for isolation using the TAG label is provided in the Examples.

25 Some desired proteins may require additions or modifications of the above-described approach as known to one of ordinary skill in the art. The method is scaleable from the laboratory bench to pilot and production facility largely because the techniques applied are well documented in each of these settings.

It is believed that a typical chicken egg produced by a transgenic animal of the
30 present invention will contain at least 0.001 mg, from about 0.001 to 1.0 mg, or from about 0.001 to 100.0 mg of exogenous protein, peptide or polypeptide, in addition to the normal constituents of egg white (or possibly replacing a small fraction of the latter).

One of skill in the art will recognize that after biological expression or purification, the desired proteins, fragments thereof and peptides may possess a conformation substantially different than the native conformations of the proteins, fragments thereof and peptides. In this case, it is often necessary to denature and 5 reduce protein and then to cause the protein to re-fold into the preferred conformation. Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art.

Production of Multimeric Proteins in Milk

In addition to methods of producing eggs containing transgenic proteins or 10 peptides, the present invention encompasses methods for the production of milk containing transgenic proteins or peptides. These methods include the administration of a transposon-based vector described above to a mammal. In one embodiment, the transposon-based vector contains a transposase operably-linked to a constitutive promoter and a gene of interest operably-linked to mammary specific promoter. 15 Genes of interest can include, but are not limited to antiviral and antibacterial proteins and immunoglobulins.

The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, 20 modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.

EXAMPLE 1

25 *Preparation of Transposon-Based Vector pTnMod*

A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given below as SEQ ID NO:57. The vector of SEQ ID NO:57, termed pTnMod, was constructed and its sequence verified.

This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak 30 sequence (ACCATG) (SEQ ID NO:5) was added to the promoter. The nucleotide in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic

polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic polyA insures proper termination without read through to potential genes downstream.

The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCATGTGTGAACTTGATATTTACATGATTCTCTTACC 3' (SEQ ID NO:91) and Altered transposase- High efficiency reverse primer (Her) 5' GATTGATCATTATCATAATTCCCCAAAGCGTAACC 3' (SEQ ID NO:92, a reverse complement primer). In the 5' forward primer ATS-Hef, the sequence CTCGAG (SEQ ID NO:93) is the recognition site for the restriction enzyme Xho I, which permits directional cloning of the amplified gene. The sequence ACCATG (SEQ ID NO:5) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:92) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site, TGATCA (SEQ ID NO:94), to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe™ PCR System (which includes enzyme, buffers, dNTP's, MgCl₂ and PCR Enhancer; Epicentre Technologies, Madison, WI). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. A band corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, CA). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

Plasmid gWhiz (Gene Therapy Systems, San Diego, CA) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatible with Xho I and Bcl I, but destroy the restriction sites. Digested gWhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting 5 the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, CA) was used and the ligation set up according to the manufacturer's protocol. Ligated product was 10 transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies, Carlsbad, CA) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT# 15544-042) medium for 1 hour at 37° C before being spread to LB (Luria-Bertani media (broth or agar)) plates supplemented with 100 µg/ml ampicillin (LB/amp plates). 15 These plates were incubated overnight at 37° C and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp) 20 were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Qiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Qiagen, Inc., Chatsworth, CA). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes and no further changes or mutations occurred due to PCR amplification. For 25 sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer Model 377 Automated Sequencer.

Once a clone was identified that contained the desired mATS in the correct orientation, primers CMVf-NgoM IV (5' TTGCCGGCATCAGATTGGCTAT (SEQ 30 ID NO:95); underlined bases denote NgoM IV recognition site) and Syn-polyA-BstE II (5' AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA (SEQ ID NO:96); underlined bases denote BstE II recognition site) were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the transposon

in pTnLac. The PCR was conducted with FailSafe™ as described above, purified using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

5 Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a
10 Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA were ligated together using a Stratagene T4 Ligase Kit and transformed into *E. coli* as described above.

15 Colonies resulting from this transformation were screened (mini-preps) as described above and clones that were the correct size were verified by DNA sequence analysis as described above. The vector was given the name pTnMod (SEQ ID NO:57) and includes the following components:

Base pairs 1-130 are a remainder of F1(-) on from pBluescriptII sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptII sk(-).

20 Base pairs 131 - 132 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 133 -1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

25 Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 1780 - 2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T
30 when such a substitution would not alter the amino acid that was encoded.

Base pairs 2988-2993 are two engineered stop codons.

Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 2995 - 3410 are a synthetic polyA sequence taken from the pGWiz vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

Base pairs 3415 - 3718 are non-coding DNA that is residual from vector pNK2859.

5 Base pairs 3719 - 3761 are non-coding λ DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 bp of the left insertion sequence recognized by the transposon Tn10.

Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

10 Base pairs 3838 - 4527 are the multiple cloning site from pBluescriptII sk(20), corresponding to bp 924-235 of pBluescriptII sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites used in constructing the vector.

15 Base pairs 4533 - 4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

Base pairs 4603 - 4644 are non-coding λ DNA that is residual from pNK2859.

Base pairs 4645 - 5488 are non-coding DNA that is residual from pNK2859.

20 Base pairs 5489 - 7689 are from the pBluescriptII sk(-) base vector - (Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptII sk(-).

Completing pTnMod is a pBlueScript backbone that contains a colE I origin of replication and an antibiotic resistance marker (ampicillin).

25 It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

30 All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA was recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, CA) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500 μ L of PCR-grade water and stored at -20°C until used.

EXAMPLE 2

Transposon-Based Vector pTnMCS

Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and 5 its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:56. The pTnMCS vector contains an avian optimized polyA sequence operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 75 nucleotides that precede the A nucleotide string.

10 Bp 1 – 130 Remainder of F1 (-) ori of pBluescriptII sk(-) (Stratagene) bp1-130
Bp 133 – 1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy
Systems) bp 229-1873
Bp 1783 – 2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316
Bp 2992 – 3344 Non coding DNA from vector pNK2859
15 Bp 3345 – 3387 Lambda DNA from pNK2859
Bp 3388 – 3457 70 bp of IS10 left from Tn10
Bp 3464 – 3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp
924-718
Bp 3671 - 3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site
20 thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp
717-673
Bp 3716 – 4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp
672-235
Bp 4159 - 4228 70 bp of IS10 right from Tn10
25 Bp 4229 - 4270 Lambda DNA from pNK2859
Bp 4271 – 5114 Non-coding DNA from pNK2859
Bp 5115 - 7315 pBluescript sk (-) base vector (Stratagene, Inc.) bp 761-2961.

EXAMPLE 3

30 *Production of Antibody in Egg White*

A transposon-based vector containing a CMV promoter/cecropin prepro/antibody heavy chain/cecropin pro/Antibody light chain/conalbumin poly A (SEQ ID NO:97) was injected into the oviduct of quail and chickens. A total of 20 birds were injected (10 chickens and 10 quail) and eggs were harvested from the birds

once the eggs were laid. Partially purified egg white protein (EW) was then run on a gel under both reducing and non-reducing conditions. Figure 5 is a picture of the gel. Lanes 1 & 18: molecular weight markers, Lanes 2 and 3: EW #1, non-reduced, reduced, respectively; Lanes 4 and 5: EW #2, non-reduced, reduced, respectively, 5 Lanes 6 and 7: EW #3, non-reduced, reduced, respectively, Lanes 8 and 9: EW #4, non-reduced, reduced, respectively; Lanes 10 and 11: EW #5, non-reduced, reduced, respectively; Lanes 12 and 13: EW #6, non-reduced, reduced, respectively; Lanes 14 and 15: EW #7, non-reduced, reduced, respectively; and Lanes 16 and 17: EW #8 Control, non-reduced, reduced, respectively. Based upon the gel results, the 10 possibility that the egg white in the treated chicken and quail contains antibody produced by the above-mentioned transposon-based vector cannot be excluded.

EXAMPLE 4

Additional Transposon-Based Vectors for Administration to an Animal

The following example provides a description of various transposon-based 15 vectors of the present invention and several constructs for insertion into the transposon-based vectors of the present invention. These examples are not meant to be limiting in any way. The constructs for insertion into a transposon-based vector are provided in a cloning vector pTnMCS or pTnMod, both described above.

pTnMOD (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:97)

20 Bp 1-4090 from vector pTnMod, bp 1 - 4090
Bp 4096-5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by 25 Mark Glassy (Shantha West, Inc)
Bp 7288-7302 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)
Bp 7309-7953 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
30 Bp 7960-8372 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
Bp 8374-11973 from cloning vector pTnMod, bp 4091-7690

pTnMCS (CHOVep-prepro-HCPro-CPA) (SEQ ID NO:98)

Bp 1-3715 from vector pTnMCS, bp 1-3715
Bp 3721-4395 Chicken Ovalbumin enhancer taken from GenBank accession #
5 S82527.1, bp 1-675
Bp 4402-5738 Chicken Ovalbumin promoter taken from GenBank accession #
J00899-M24999, bp 1-1336
Bp 5745-5915 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
Bp 5922-7286 Heavy Chain gene construct taken from antibody RM2 provided by
10 Mark Glassy (Shantha West, Inc)
Bp 7287-7298 Pro taken from GenBank accession # X07404, bp 719-730 (does not
include Lysine)
Bp 7305-7949 Light Chain gene construct taken from antibody RM2 provided by
Mark Glassy (Shantha West, Inc)
15 Bp 7956-8363 Conalbumin polyA taken from GenBank accession # Y00407, bp
10651-11058
Bp 8365-11964 from cloning vector pTnMCS, bp 3716-7315

pTnMCS(CHOvep-prepro-HCPro-Lys-CPA) (SEQ ID NO:99)

20 Bp 1 - 3715 from vector pTnMCS, bp 1-3715
Bp 3721 - 4395 Chicken Ovalbumin enhancer taken from GenBank accession #
S82527.1, bp 1-675
Bp 4402 - 5738 Chicken Ovalbumin promoter taken from GenBank accession #
J00899-M24999, bp 1-1336
25 Bp 5745 - 5915 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
Bp 5922 - 7286 Heavy Chain gene construct taken from antibody RM2 provided by
Mark Glassy (Shantha West, Inc)
Bp 7287 - 7301 Pro taken from GenBank accession # X07404, bp 719-733 (includes
Lysine)
30 Bp 7308 - 7952 Light Chain gene construct taken from antibody RM2 provided by
Mark Glassy (Shantha West, Inc)
Bp 7959 - 8366 Conalbumin polyA taken from GenBank accession # Y00407, bp
10651-11058

Bp 8368 – 11967 from cloning vector pTnMCS, bp 3716-7315

pTnMCS (CMV-prepro-HCPro-CPA) (SEQ ID NO:100)

Bp 1 – 3715 from vector pTnMCS, bp 1-3715

5 Bp 3721 – 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 5548 - 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

10 Bp 6913 - 6924 Pro taken from GenBank accession # X07404, bp 719-730 (does not Lysine)

Bp 6931 – 7575 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7582 – 7989 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 7991 – 11590 from cloning vector pTnMCS, bp 3716-7315

pTnMCS (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:101)

Bp 1 – 3715 from vector pTnMCS, bp 1-3715

20 Bp 3721 – 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 5548 - 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

25 Bp 6913 - 6927 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 6934 – 7578 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7585 – 7992 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 7994 – 11593 from cloning vector pTnMCS, bp 3716-7315

pTnMod (CHOvep-prepro-HCPro-CPA) (SEQ ID NO:102)

Bp 1-4090 from vector pTnMod, bp 1-4090

Bp 4096-4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675

5 Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

10 Bp 7662-7673 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

Bp 7680-8324 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8331-8738 Conalbumin polyA taken from GenBank accession # Y00407, bp 15 10651-11058

Bp 8740-12339 from cloning vector pTnMod, bp 3716-7315

pTnMod (CHOvep-prepro-HCPro-LYS-CPA) (SEQ ID NO:103)

Bp 1-4090 from vector pTnMod, bp 1-4090

20 Bp 4096-4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675

Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

25 Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7662-7676 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7683-8327 Light Chain gene construct taken from antibody RM2 provided by 30 Mark Glassy (Shantha West, Inc)

Bp 8334-8741 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8743-12342 from cloning vector pTnMod, bp 3716-7315

pTnMod (CMV-prepro-HCPro-CPA) (SEQ ID NO:104)

Bp 1-4090 from vector pTnMod, bp 1-4090
Bp 4096-5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy
5 systems), bp 230-1864
Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by
Mark Glassy (Shantha West, Inc)
Bp 7288-7299 Pro taken from GenBank accession # X07404, bp 719-730 (does not
10 include Lysine)
Bp 7306-7950 Light Chain gene construct taken from antibody RM2 provided by
Mark Glassy (Shantha West, Inc)
Bp 7557-7969 Conalbumin polyA taken from GenBank accession # Y00407, bp
10651-11058
15 Bp 7971-11970 from cloning vector pTnMod, bp 3716-7315

All patents, publications and abstracts cited above are incorporated herein by
reference in their entirety. It should be understood that the foregoing relates only to
preferred embodiments of the present invention and that numerous modifications or
20 alterations may be made therein without departing from the spirit and the scope of the
present invention as defined in the following claims.

CLAIMS

We claim:

1. An isolated polynucleotide comprising two or more genes of interest and two or more pro nucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence and each of the two or more genes of interest may be the same or different.
5
2. The polynucleotide of claim 1, wherein a most 5' pro nucleotide sequence of the two or more pro nucleotide sequences is a part of a prepro nucleotide sequence.
10
3. The polynucleotide of claim 1, wherein two genes of interest and two pro nucleotide sequences are arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.
15
4. The polynucleotide of claim 3, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence and the pro nucleotide sequence is a cecropin pro sequence.
20
5. The polynucleotide of claim 3, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4 and the pro nucleotide sequence comprises a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.
25
6. The polynucleotide of claim 1, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.
30
7. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or more genes of interest, wherein each gene of interest encodes a part of the multimeric protein, each gene of interest is operably-linked to a pro nucleotide sequence, and each of the two or more genes of interest may be the same or different.
35
8. The method of claim 7, wherein the multimeric protein is an associated multimeric protein.

9. The method of claim 7, wherein the multimeric protein is a multivalent multimeric protein.
- 5 10. The method of claim 7, wherein a most 5' pro nucleotide sequence of the two or more pro sequences is a part of a prepro nucleotide sequence.
- 10 11. The method of claim 7, wherein the polynucleotide comprises two genes of interest and two pro nucleotide sequences arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.
- 15 12. The method of claim 7, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.
- 20 13. A method of producing a protein in an individual comprising administering to the individual a polynucleotide comprising a cecropin prepro nucleotide sequence operably-linked to one or more genes of interest, each gene of interest encoding the multimeric protein.
14. The method of claim 13, wherein a first gene of interest is an antibody heavy chain and a second gene of interest is an antibody light chain.
- 25 15. The method of claim 13, wherein the protein is a multimeric protein and the cecropin prepro nucleotide sequence is operably-linked to two or more genes of interest, wherein each gene of interest encodes a part of the multimeric protein.
- 30 16. The method of claim 15, wherein the multimeric protein is an associated multimeric protein.
17. The method of claim 15, wherein the multimeric protein is a multivalent multimeric protein.
- 35 18. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or more genes of interest, wherein each gene of interest encodes a part of the

multimeric protein and wherein each gene of interest is operably linked to a gene encoding for a cleavage site.

19. The method of claim 18, wherein a transposon-based vector comprises the polynucleotide and further comprises a transposase gene operably linked to a first promoter and wherein;

- 5 a) the first promoter comprises a modified Kozak sequence comprising ACCATG;
- 10 b) the two or more genes of interest are each operably-linked to one or more additional promoters; and,
- c) the two or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

15 20. The method of claim 18, wherein a transposon-based vector comprises the polynucleotide and further comprises a transposase gene operably linked to a first promoter and an avian optimized polyA sequence, and wherein;

- 20 a) the two or more genes of interest are each operably-linked to one or more additional promoters; and,
- b) the two or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

25 21. The method of claim 18, wherein the multimeric protein is an associated multimeric protein.

30 22. The method of claim 18, wherein the multimeric protein is a multivalent multimeric protein.

23. The method of claim 18, wherein the polynucleotide further comprises a cleavage site.

35 24. An animal comprising the isolated polynucleotide of claim 1.

25. The animal of claim 24, wherein the animal is a bird.

26. An egg produced by the animal of claim 25.

27. The egg of claim 26, wherein the egg comprises a multimeric protein encoded by the isolated polynucleotide.

28. The animal of claim 24, wherein the animal is a mammal.

5

29. Milk produced by the mammal of claim 28.

30. The milk of claim 29, wherein the milk comprises a multimeric protein encoded by the isolated polynucleotide.

10

31. A method of producing a multimeric protein comprising:

a) administering to an egg-laying animal a composition comprising the polynucleotide of claim 1 in an acceptable carrier; and,

b) permitting the one or more genes of interest to be expressed

15 into the multimeric protein.

32. The method of claim 31, further comprising

a) collecting an egg from the egg-laying animal;

b) harvesting egg white comprising the multimeric protein; and,

20 c) purifying the multimeric protein.

33. The method of claim 31, wherein the egg-laying animal is a bird.

34. A method of producing a multimeric protein comprising:

25 a) administering to an intramammary duct system of a mammal a composition comprising the polynucleotide of claim 1 in an acceptable carrier, and,

b) permitting the one or more genes of interest to be expressed into the multimeric protein.

30

35. The method of claim 34, further comprising

a) collecting milk from the mammal, wherein the milk comprises the multimeric protein;

b) purifying the multimeric protein.

35

36. The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

37. The polynucleotide of any of the preceding claims, wherein the two or more pro nucleotide sequences each comprise a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.

5

38. The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence.

10 39. The polynucleotide of any of the preceding claims, wherein the pro nucleotide sequence is a cecropin pro nucleotide sequence.

15

1/5

FIGURE 1

Prom	prepro	Gene of Interest	pro	Gene of interest	polyA

2/5

FIGURE 2

IS	Oval Prom	prepro	Heavy chain	pro	Light chain	polyA	IS
----	-----------	--------	-------------	-----	-------------	-------	----

FIGURE 3

Prom	Cecropin prepro	Gene of Interest	CS	Gene of Interest	PolyA

4/5

FIGURE 4

IS	Prom	SS	Gene of Interest	CS	Gene of Interest	polyA	IS

5/5

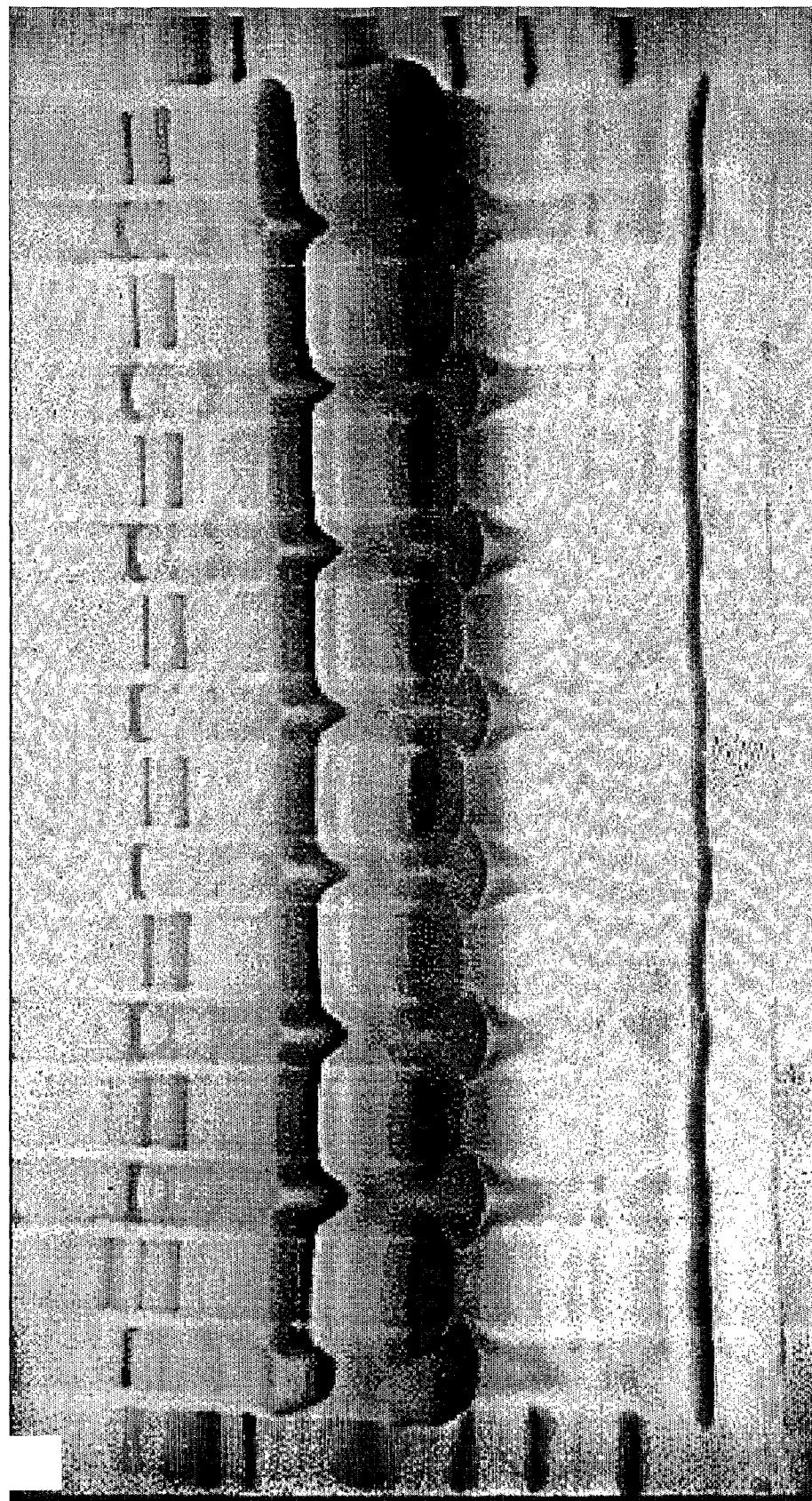


FIGURE 5

Appendix A

5 SEQ ID NO:1 (cecropin pro)
 GCG CCA GAG CCG AAA

10 SEQ ID NO:2 (cecropin pro extended)
 GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG

10 SEQ ID NO:3 (cecropin prepro)
 AAT TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA
 ACA GTT TCG GCT GCG CCA GAG CCG AAA

15 SEQ ID NO:4 (cecropin prepro extended)
 AAT TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA
 ACA GTT TCG GCT GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG

20 SEQ ID NO:5 (modified Kozak sequence)
 ACCATG

25 SEQ ID NO:6
 atg ctg ggc atc tgg acc ctc cta cct ctg gtt ctt acg tct gtt gct aga
 tta

30 SEQ ID NO:7
 atg att cct gcc aga ttt gcc ggg gtg ctg ctt gct ctg gcc ctc att ttg
 cca ggg acc ctt tgt

30 SEQ ID NO:8
 atgg gcagagcaat ggtggccagg ctggggctgg ggctgtgtct gctggcactg
 ctcctaccca cgcagattta ttcc

35 SEQ ID NO:9
 atgaatctat cgaacatttc tgaggtaaaa gtattaacac tggtggttag cgctgccatc gct

40 SEQ ID NO:10

40 atgaccatcc tttccttac tatggttatc tcaatctca gttgcattaa agctgccccg
 atgaaaagaag ctatgttaag aggacatggc agcttggctt acccaggctt tcggaccac
 gggactcttggaaagcataac tggggccaaat gctggttcaa gaggactgac atcaactggcg
 gacacttttggaaacacgtat agaggagttt ctatgttaag atcaggacat ccagccccagt
 gaggaaaaaca aggtatggaa cttgtacaca tcccgagtca tgctgagcag tcaagtgcct
 ttggaaacccccactgtctt tctgtctcgag gaggacaaaa actaccttggatgctgcaaac
 atgtccatga gatccggcg tcaactgtac

45 SEQ ID NO:11
 cgttttttc tcttatcttt tctcgctttc gctttttct cgtcgccat tgctttctcc
 gacgacgatc cggtgatccg acaagttgtatc tggggaaacg atgacaacca tatgttaaac
 gccgagcatacttttcaacttttt

50 SEQ ID NO:12
 atg tccatcttgtt tttatgtat atttcttgca tatcttcgtg gcattcagtc aactaatatg
 gatcaaagga gtttgcaga agatcaatg aattctctca ttatcaaact cattcggcg
 gacatcttgc aaaaacaaatgt ttctaaatgc gttatggatg tcaaggaaaa ctatcaaaac
 atatgtgcaga aatgtatggaa ccaccaggat atggatggatg atgaaaatgt gaaatcagac
 ttggggggatg ttatcttcaat ggtatcagac ctcctaaatgc agcagagacg ctacaactct
 tc tccttaatgtatc caacacacca ctggaaaccac caccactgtatc cttcacatgg
 tg gaagttcagt ggtatcaaac agaacatcttc gaaggaaaaag gt

SEQ ID NO:13
 atgggtgcata tgactcctga ggagaagtc ggcgttactg ccctgtgggg caagggtgaac
 gtggatgaag ttgggtgtga ggccctgggc aggctgtgg tggtctaccc ttggacccag
 5 aggttctttg agtcctttgg ggatctgtcc actcctgtatc ctgttatggg caaccctaag
 gtgaaggctc atggcaagaa agtgtctgg gcctttagtg atggctggc tcacctggac
 aacctaagg gcacctttgc cacactgagt gagctgcact gtgacaagct gcacgtggat
 cctgagaact tcaggctcct gggcaacgtg ctgggtctgtg tgctggccca tcactttggc
 aaagaattca cccccaccagt gcaggctgcc tatcagaaag ttgtggctgg tggcataat
 10 gcccctggccc acaagtatca ctaagctcgc ttcttgcgtg tccaatttct attaaagggtt
 cctttgttcc ctaagtccaa ctactaaact gggggatatt atgaagggcc ttgagcatct
 ggattctgcc taataaaaaa catttatttt cattgc

15 SEQ ID NO:14
 gcatggggac ggcgttctc cagcgcgggg gctgtttct cctgtgcctt tcgctgtgc
 tcctggctg ctgggcggag ctgggcagcggg

20 SEQ ID NO:15
 20 cggaaacgatt caaaacctct ttactgcccgt tatttgctgg atttttattt ctgttttattt
 tggttctggc aggac

25 SEQ ID NO:16
 25 ggaggtctgg ggaggcttag tgcagcctgg agagtccctg aaactctcct gtgaatccaa
 tgaatacgaa ttcccttccc atgacatgtc ttgggtccgc aagactccgg agaaggaggct
 ggaggtggtc gcagccatta atagtgtatgg tggtagcacc tactatccag acaccatgg
 gagacgatcc atcatctccaa gagacaatac caagaagacc ctgtacctgc aaatgagcag
 tctgaggctc gaggacacag cttgtatata ctgtgcaga cacacgatga gcaaaagtta
 30 ctgtgagctc aaactaaaac ctccctgcaga gcatccagga ccagcagggg ggcgcggagag
 acacagagtt gtgaaat

35 SEQ ID NO:17
 35 acatccattc ttctgtgagt ttcaactcgaa gagcagegtg tcactgcgga caagccagcc
 agcteaccat ggctggacct cccagggta cagacctctg ggaactggcc ctgagcctca
 cttcacggac acaggtgtcc cgccaaagtg ggtctcagag caacagtgtg tgcatgtgc
 gtcacatctt cctttgttctt tgcatgactg actacacccca agaagtgtgc ccctggagg
 aaagcatatt tgcaaccag atcataataa aatcagaaat gcagcaaacc tttaaaatat
 ccagacttgg

40 SEQ ID NO:18
 40 tggaagcaag agggagttatg ctaacttcat g

45 SEQ ID NO:19
 45 atcaattaca agaggg

50 SEQ ID NO:20
 50 atgaagttmg cataactccct cttgttccaa ttggcaggag tcagtgttc agtkatcaat
 tacaagaga

55 SEQ ID NO:21
 55 aattcttaat taattattgt ggtgtcacaa taacttttc

60 SEQ ID NO:22
 60 cccccccggat ccatggccgc taaattcgtc gtggttctgg ccgttgcgt cgccctgagc
 cactcggcta tggtgccggc caagaagaac ggctac

65 SEQ ID NO:23
 65 cccccccggat ccatgaaaact cctggtcgtg ttgcgttgcgt gctgtgcggc tgccagcgct

SEQ ID NO:24
cagtgtacgg cggctcgagg cagaagtcgg gacgcata

5
SEQ ID NO:25
gaattcatta tcagcacgga ccagattctt ggatcaggat agaagtctga cgtttacgg
tttcgaagca cagaacgcattcgttagcgta aagtgtcacc gtcggtagcc caaaccggac
10 ggtattccag agtgcaaccg ttcaagttcgatgttagcattt agcttcacga cccagagat
ccatggatcc cccttcgcgttcttcag ttcaagcat tgcgatttt ttaagcaacg
cactctcgat tcgttagagcc tcgttgcggtt tgttgcacg aaccatatg

SEQ ID NO:26
ttcacaggca ggaaaaatggta gagagggca tgcataatgc ctcactgtgg

15
SEQ ID NO:27
aagcttctcg tggaaaccaa cccaaattatgt tagtattgca ttctgtgtac tataatgg
aatataaaa atatttaaa atacctccat tttgtttatc cttttatgtga agatgatacc
tgcaaaagac atggctaaag ttatgtatgt catgttggca atttggatttc ttacaaaatc
20 ggtatggaaa tctgttaagt aagtactgtt ttgccttggaa attggatttt taatgttgc
tttacatccat cgaagtgggg agctaattggg aagtggccct ctctgttttcttcttccca
ggaagagat

SEQ ID NO:28
25 atggctacag gtcggccggac gtccctgtccttgc ctggattttgc gctgtctgc cctgcctgg
cttcaagagg gcagtgc

SEQ ID NO:29
atgagggtct ttgtataatct tgggtgttttgc
30 cttcctgccc ctggctgtccttgc tgggg

SEQ ID NO:30
atgc acctgagaat ccacgcgaga cggaaaccctc ctgcggggcc ggcctggac
cttggatcttgc gtcggatgttgc atcgatgtatc
35

SEQ ID NO:31
atg gccatttagtg gaggccctgt gtcaggatattt ttcatcatag ctgtgtatgat gagcgctc
gaatcatggg ctatcaaaga agaacatgttgc atcatccagg ccgagtttcta tctgaatct
gaccaatcg gcgagtttat gtttgac

40
SEQ ID NO:32
aggggggatc cccggagacc ttccggtagc aactgtcacc ttgtatgtgg cgatcctgag
ctcctactg gtcggggc

45
SEQ ID NO:33
atg gtgtgtctga ggctccctgg aggttcctgc atggcagttc tgacagtgc
actgtatggtg ctgagctccc cactggcttttgc

50
SEQ ID NO:34
1 gaacgatata aggagcgaat actactggta aactaatggaa agaaatctgc tgccaccactg
61 gatattggga gtgtgtggca tgcatttcata tcattcaggaa actctaaaaa agaaccggat
121 ggtgttagcc aaacagctgt tttgtggcga attgttagaa catcttctgg agaaggacat
181 catcaccttgc gaaatgggg agtcatccaa ggccaaatgttgc ggcagtttca gccagaatgt
241 ggaactccttgc aacttgcgtc ctaaggggg tcccccacgttgc tttgtatgcct tctgtgttgc
301 actgtatggggatc accaagcaag ggcacccgttgc ggtatgttgc ctcaccaccc tttctggc
361 tcagcatgttgc cttccacccgttgc ttagctgttgc ctacgacttgc agtctccctt ttcgggtgt
421 tgatgtatggggatc ccccttatac agaagtcggc cctgtcgaca gataactgtgg aacactccct
481 agacaataaa gatggatgttgc ttcgtatgttgc ggtatgttgc tgcactcctg aattttatca

541 aacacacttc cagctggcat atagggtgca gtctcgccct cgtggcctag cactgggttt
 601 gagcaatgtg cacttcaactg gagagaaaaga actggaaattt cgctctggag gggatgtgga
 661 ccacagactact ctatgcaccc tcttcaagct tttgggctat gacgtccatg ttctatgtga
 721 ccagactgca cagggaaatgc aagagaaaact gcagaatttt gcacagttac ctgcacaccc
 5 781 agtacacggac tcttgcacatcg tggcactctt ctgcacatgtt gtggaggggcg ccatctatgg
 841 tgtggatggg aaactgctcc agtccaaga ggttttcag ctcttgaca acgccaactg
 901 cccaaagccta cagaacaaac caaaaatgtt cttcacatccag gcctgcctg gagggtctat
 961 tggatccctt gggcacctcc ttctgttacac tgctgccacc gcctcttgc ctctatgaga
 1021 ctgatcgtgg ggttgaccaa caagatggaa agaaccacgc aggatcccct
 10 gggtgcgagg
 1081 agagtgtatgc cggtaaaagaa aagttgccga agatgagact gcccacgcgc
 tcagacatga
 1141 tatgcggcta tgcctgcctc aaaggactg ccgcacatgcg gaacacccaaa
 cgagggttct
 15 1201 ggtacatcga ggctcttgct caagtgtttt ctgagcgggc ttgtgatatgc
 cacgtggccg
 1261 acatgtctgtt taaggtgaac gcacttatca aggatcgaa aggttatgtct
 cctggacacg
 1321 aattccaccc gtgcaaggag atgtctgaat actgcagcac tctgtgccgc
 20 1381 tggccagg acacccctccc acatgtatgtc acctccccat catccacgc
 aagtggaaagc
 1441 cactggacca caggagggtgt gatagagcct ttgatcttca ggatgcacgg
 tttctgttct
 25 1501 gccccctcag ggatgtggga atctcccaga cttgtttcct gtgccttcatca
 tctctgcatt
 1561 tgagtgtggg actccaggcc agtccctttt ctgtgaagcc ctttgcctgt
 agagccacgc
 1621 ttgggttggac ctattgcacat gatatgttca gctgcagttt aagagcctga
 30 1681 caagtgaagt
 1681 tggtaaacaca gtgtggttat ggggagaggg catataaatt cccatatttt
 gtgttcagtt
 1741 ccagctttt tagatggcac ttttagtgatt gcttttatttattt cattagttaa
 gatgtctgag
 35 1801 agaccatctc ctatcttttta ttccattcat atcctccgc cttttgtcc
 tagagtgaga
 1861 gtttggaaagg tgcctaaatt taatgttagac attatctttt ggctctgaag
 aagcaacat
 1921 gactagagac gcacccctgtt gcaatgttca gaagcggcct gtgcgttccc
 40 ttccatgt
 1981 cagccaccc cagtgaaagg acactcttgg ctgcgttggg ctcaaggcac
 cgcacccctgt
 2041 cagccaaacat tgccttgcattt ttgttaccttta ttgatctttt cccatggaaag
 tctcaaaat
 45 2101 ctttcgttgg ttgtttctct gagctttgtt actgaaatga gcctcggttgg
 gagcatcaga
 2161 gaaggccagg aagaatggtg tggttccctta gactctgtaa ccacccctct
 gtctttttcc
 2221 ttccatgttcaaa acgtccatct ctctccctta ctattccac ttccattca
 tcaacccat
 2281 ttccatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt
 2341 ttcttttaccc tcaatttcattt ctgttggctg ctgccacccaa atctgtcttag
 aatccctgttt
 55 2401 tacaggatca tggtaaatgtt caaaagatgtt atgttagtttctt ttgttccctgc
 ttctcttttctt
 2461 agtattaaac ttccttttgcatttattatgtgg cttttatttcttcttcttcttctt
 atgttattgt
 2521 ttccatgttcaaa accttta tccctgcttta tctgaaactt cccaaacttcc
 ctgttcccttta

2581 agactttttt tttttttttt tttttttttt agacagagtc tcgctctgtc
 gcccaggctg
 2641 gagggcagtgc acacgatctc agctcactgc aagctccaac tcccggttca
 acggcattct
 5 2701 cctgcctcag ctttccaagt agctggact acagggtccc gccaccgtgc
 ccggcttaatt
 2761 tttttgtatt ttttagtagag acagggtttc accatgttag ccgggatgg
 ctgtatctcc
 2821 tgacctcatg atccacccac ctcagcctcc caaagtgttggattacagg
 10 cgtgaggccac
 2881 tgcgccccggg caagaccctt ttttaaaaaaa aaaaaaaaaaa aaacttccat
 tcttttttcc
 2941 tccagtcgt tctcacataa cagagtagttt ttgggtttta atttttttt
 gttgtttgt
 15 3001 gttttttgtt ttttaaggtg agttctcaact atgtttctca gactggcttc
 gaactcttgg
 3061 cctcaagcca tttcccgcc tcagcctctc aaatagctgg gcttacaggc
 atgaggccac
 3121 acacctggcc aggatttggg tttttaaata taaatctgat cacccccctg
 20 cttagaaccc
 3181 ttctgttttc tattaccctt cattttaaat gtaaactctt caccttgggtt
 tatgagaact
 3241 ggttcttgcc ttcccttga acctcattaa atggtgattt cttgtcaagc
 tccagcccgaa
 25 3301 gttttttttt ctcagcttctt aattttgtgc tctttctgc ctttttctg
 ggccttctca
 3361 gtcctccacc cccaccactc ttgactcagg tgggtttttt cttctcaag
 tcttgacaat
 3421 tccccggggcc ttcagtcctt gaggactcta cttctgtgtc tgcaccaca
 30 tcttgcttt
 3481 tccccctcatt gcatttattt cagtttataat atatgctact tttacttgg
 cattttctgtc
 3541 tccccctacca ggctgttaat gagggcagaa accttgggttgc ttttattcac
 catcatgtac
 35 3601 caagtgtttt gcacatagtg ggccttcattt aaatgtttgt tgaataaaag
 agggaaagaag
 3661 gcaagccaaac cttagctaca atcctacctt ttgataaaat gttcccttttgc
 acaatataca
 3721 cggatttattt tttgtacttt gtttttcattt gtgttttgc tttatccact
 40 ggcattttt
 3781 gtccttgaa gacatatcat gtgtgagata acttccttca catctcccat
 ggtcccttgc
 3841 aaaatgcttag gcctgttagta gtcaagggtgc tcaataaaata tttgtttgg
 tggtttgc
 45 3901 gccttgctgc caagtcttcac ctttgggtcg acatagtagt gaagtatttgc
 agagagagaa
 3961 ctttccactt cccactgcca ggattttgttta ttgccatcggttgc
 aatgctcata
 4021 ttttattttttt aaaaaaaaaaaa aaaaa
 50

SEQ ID NO:35

1 tccagatcat ctgttctcac caccaaggcc atgggtgtctt cagcgactat ctgcagtttgc
 61 ctacttctca gcatgctctg gatggacatg gccatggcag gttccagctt ctgtggccca
 55 121 gagcaccaga aagcccgca gagaaggaa tccaaagaagc caccagctaa actgcagccca
 181 cggatctgg aaggctggc ccaaccagag gacagaggac aagcagaaga ggcagaggag
 241 gagctggaaa tcaggttcaat tgccttcac gatgttgc tcaagctgtc agggatctca
 301 taccagcagc atggccgggc ctttggaaaat ttttttgc ttttttgc
 361 aaagaggcgca cagctaaata gtaaccactg acaggactgg tccctgtact ttcctcttca
 421 gcaagaactc acatccagct tctgccttctt ctgcaactcc cagactctc ctgctgactt
 481 acaaataaaat gttcaagctt t

SEQ ID NO:36
 5 atgaagctgc ttgcaatggt tgcactgctg gtcaccatct gtagcctaga aggagcttg
 gttcgagac

SEQ ID NO:37
 10 atgaagctgc ttgcaatggt tgcactgctg gtcaccatct gtagcctaga aggagcttg
 gttcgagac

SEQ ID NO:38
 15 atggcctt gccaacggct cgaccctgt tggggtaatg tgggacccccc gcccctggca
 gcctcctgtt cctgcttcc agcctcgat gggtcagcc c

SEQ ID NO:39
 20 atgcccgcgc tggcttccta cctccttaggt gtctggctgc tcctgagcca acttcccaga
 gaaatcccag gc

SEQ ID NO:40
 25 atgacagc atcaattgtc gtttaccat cgctttgggtt aatattaatt atttttactg
 caccctatac tcactgt

SEQ ID NO:41
 30 atgaaagtc ctgcttggc acctgctgtc gtcagtctc ttctccagtg tgttcagcag
 ttgtcagagg gactgtctca catgccagga gaagctccac ccagccctgg acagcttcga
 cctggaggtg tgcattctcg agtgcgaaga gaaggtcttc cccagccccc tctggactcc
 atgcaccaag gtcatggccca ggagcttttgc agcgttcgc cttggccccc cagagcatgt
 ggcggctgtc ctctaccaggc cgagagcttc ggagatgcag catctgcggc gaatccccgg
 agtccggcgc ttgttccagg agcaggaaga gcccggccct ggcattggagg aggctgggtga
 gatggagcag aagcagctgc agaagat

SEQ ID NO:42
 35 aattcatgaa gttgggttact ttcatcttt tgttgttctt gttctttctt gcttactcta
 gaggtttttt cagacg

SEQ ID NO:43
 40 atgaagtggg ttactttcat ctcttttttg ttctgttctt ctctgttta ctcttagaggt
 gtttccagac gc

SEQ ID NO:44
 45 1 gaattctcaa tggcaaaggc aagtgtacat tataaatagc aaaacagctg gcttggacca
 61 ttttgcggc cagtcaccca gttgagggtt ttgaatgaca tcataaccct caagagggtt
 121 ttgtctagcca gctgggttta tttagataac aaaaaaatca gagaagaaaa acacactctg
 181 gcacacagac tccctctgtc atacacacac acacacacac acacacacac acacacacac
 241 agaggtttga gtttatatggaa aaattcaaac aacaggaaaa ttgtttggccccc cccaggttacc
 301 cttctccca agtgggtgggg tggggagggg acagtgcacag gcagcttagt agaagaataaa
 361 agaaaaatgt tctatttcag ttgggtttta cagctggca tagtcttgc ctcattcgcag
 421 gagaaaaatgt atgagacagt gcccataagg gaccaatcca atgctgcctg cccctccata
 481 ctttcttagga aatgagatca cacccttcac ttggcaactg ggacaaggggg tcacccgagt
 541 gctgtcttcc aatctacttt accccagtc ttccagggtt aaaattgttag agtggctgg
 601 agaggggtctt atcgtctttt ttgtttttaa taatgcattt gctctagaat
 661 ctaaaatttgc tctccatcc cccatattcc ttaataactg gtaaggtgtt ttagcagacg
 721 tttgtgtctt catggccaggc agaaaggtaa tcagaaaac gatcttatt ttctatggca
 781 gcataagtat tttaatgtt gcaaccctg tcgttaacac acatttttt aaggggaaaaaa
 841 aatgcttctg tgctctagtt ttaaaatgca aaggatgtat gtttttttgc accatgccc
 901 aaaaagtcc tactcaataa cttggccaga agagggagag agagagaagg caaatgttcc
 961 cccagctgtt tcctgtctac agtgcgtgtt tttgttagat aatgtgagg attttgtgtt

1021 aatcccttctt ctgttgctta aatctcaactg tcactgctaa attcagagca
 gatagagcct
 1081 ggcgaatggataaaatgcct caaaaattgaa atgtgacatt gctctcaaca
 tctccatct
 5 1141 ctctggattt cttttgtttt cattattcct gctaaaccaat tcattttcag
 actttgtact
 1201 tcagaagcaa tggggaaaaat cagcagtctt ccaacccaaat tatttaagt
 ctgctttgt
 1261 gatttcttga aggtaaatat ttcttactct ttgaagtcat tggggaaattc
 10 ggatcccact
 1321 gtaataatag catcttcat ttccgttagta aacgtttcta gatattttgt
 ctcaattcat
 1381 tggaaatagga acccataaaag aaagggggttc agggaggact cctccaaaga
 tccacagtag
 15 1441 ccaggggaaat aaacacaggt tggtggatgc cgagacacgc tccatccaca
 actccctgt
 1501 ggggttctcat gtactctatt ggcttctgtg ctggtagtc ctgattaatg
 acagtctgtgg
 1561 aatcggtggaa gtcaatgcac ttctgtccca ccccaactccc cttgcaagga
 20 tcaaggagga
 1621 aacctgaacc tccctctgtt tcttggcag gtgaagatgc acaccatgtc
 ctccctcgcat
 1681 ctcttctacc tggcgctgtg cctgctcacc ttcaccagct ctgccacggc t
 25

SEQ ID NO:45
 30 atgaag ccaattcaaa aactcttagc tggccttatt ctactgactt cgtgcgtggaa
 aggctgtcc agccagcact ggctctatgg actgcccctt ggaggaaaga gagatgccga
 aaattt

SEQ ID NO:46
 35 atgagatt tccttcaatt tttactgcag ttttattcgc agcatcctcc gcattagctg
 ctccagtcaa cactacaaca gaagataaa cggcacaaat tccggctgaa gctgtcatcg
 gttacttaga tttagaaggg gatttcgtatg ttgtgtttt gccatttcc aacagcacaa
 ataacgggtt attgtttata aatactacta ttgccagcat tgctgctaaa gaagaagggg
 tatcttggaa taaaagagag gctgaagctt

40

SEQ ID NO:47
 atgaa gtgggttaacc tttatttccc ttcttttctt cttagctcg gcttatttcca
 ggggtgtgtt tcgtcgaga

45 SEQ ID NO:48
 1 atgaagatgg ttccttcctc ggcctccgc tgcctctcg tgctctgtct gtcctgacc
 61 gctccatca gctgctctt cggccggacag agagactcga aactccgcct gctgtgcac
 121 cggtacccgc tgcaggggctc caaaccggacatgactcgct cgccttggc cgagctgctc
 181 ctgtcgacc tccctgcagggg ggagaacggag gctctggagg aggagaactt ccctctggcc
 50 241 aaggaggac cccggggacgc ccacccgcctc ctagagccgg cccggccgg gggccctctg
 301 ctgcggccccc gggagagaaa gggccggctgc aagaacttct tctgaaaaac ctccaccc
 361 tgctga

SEQ ID NO:49
 55 atggccgggc gaggggggcag cggcgctgtg gctctgtgc gggcactggc tgcctgggg
 tggctctgg ggcggcgaagc ccaggagccc gggccggcc cggccggcat gaggccggc
 cggccggctgc agcaagagga cggcatctcc ttctggatacc accgctacc cggatgcgc
 gaggccgtcg tgcctgtgt gctgcagtgc accgcccata gcaaggattt aacgggtgggg
 cgcaggctcg aaggccggga gctctggc atcgagctgt cccgacaaccc tggcgatccat
 60 gaggctgggtg agcctgaatt taaatacatt gggaaatatgc atggaaatga ggctgtggaa

121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgatac catatcataa
181 tatgtacatt tatattggct catgtccaaac attaccgca tggacattt gattattgac
241 tagttatcaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttcg
301 cgttacataa cttacggtaa atggcccgcc tggctgaccc cccaaacgacc cccgcatt
361 gacgtcaata atgacgtatg tccatagt aacgccaata gggacttcc attgacgtca
421 atgggtggag tatttacggt aaactgccc cttggcaga catcaagtg atcatatgc
481 aagtacgccc cttatggcg tcaatgacgg taaatggccc gcctggcatt atgcccaga
541 catgacccta tgggacttgc tcaatggca gtacatctac gtatagtc tgcattttac
601 catggtgatg cggtttggc agtacatcaa tgggctgga tagcggttg actcacgggg
661 atttccaagt ctccacccca ttgacgtca tgggagttt tttggcacc aaaatcaacg
721 ggactttcca aatgtcgta acaactccgc cccattgacg caaatggcg gtaggcgtgt
781 acgggtggag gtctatataa gcagagctcg ttagtgaat cgtcagatcg cttggagacg
841 ccatccacgc tggttggacc tccatagaag acaccgggac cgatccagcc tccggcccg
901 ggaacgggtc atggaaacgc ggattccccc tgccaaagat gacgtaaat cccctatag
961 actctatagg cacacccctt tggctttat goatgctata ctgtttttgg cttggggct
1021 atacacccccc gcttcttat gctatagggt atggatagc ttgcctata ggtgtgggtt
1081 attgaccatt attgaccact cccctattgg tgacgataat ttccattact aatccataac
1141 atggctctt gccacaacta tctctattgg ctatatgcca atactctgtc cttagagac
1201 tgacacggac tctgtattt tacaggatgg ggtcccat attatttaca aattcacata
1261 tacaacaaacg cggcccccc tggcccgagt ttttattaaa catagcgtgg gatctccacg
1321 cgaatctcg gtacgttgc cggacatggg ctcttctcg gtacgcccgg agcttccaca
1381 tccgagccct gttcccatgc ctccagggc tcatggcgc tggcagctc ctgtctctt
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gecgcacaag
1501 gccgtggcg tagggtatgt gtctaaaaat gagcgtggag attgggctcg cacggctgac
1561 gcagatggaa gacttaaggc agcggcagaa gaagatcgag cagctgtgt tttttatttgc
1621 tgataaagggt cagaggtaaatcccggtcgcc gttgtttaa cggtggaggg cagttgtatc
1681 tgagcgtac tctgtgtgc cggccggcc accagacata atagctgaca gactaacaga
1741 ctgttccctt ccatgggtt tttctgcgtt aaccgtcgga ccatgtcgaa atcgtatc
1801 ttacacgact ctcttacca atttgcctt gaaatttacaact taaaacgact caacagctt
1861 acgttggctt gcccacgcatt acttgcgtgt aaaactctca ctcttacca gatggccgt
1921 aacctgccaa cccaaagcgag aacaaaacat aacatcaac gaatcgaccg attgttaggt
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgttag ctttatctgt
2041 tccggcaata cgatggccat tggacttgc gactggctcg atattcgta gcaaaaaacgca
2101 cttaggtat tgcgagcttc agtgcacta cacgggtcggtt ctgttactct ttatggatgg
2161 gcttcccgcc tttcagagca atgttcaaaag aacatctcgat accaataatctt accatcgat
2221 gcgagattc taccggatca caccacaccg ctatgtcgat ttttttttttgcgtt
2281 ccatggtata aatccgttgc gaagctgggt tggactgtgtt ttttttttttgcgtt
2341 gtacaatatg cagacctagg agcggaaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggtat aagaggctga ctaaaagcaa tccaatctca
2461 tgcacaaattt tattgtataa attcgcctt aaggccggaa aaaatcagcg ctcgacacgg
2521 actcatgtc accacccgc tccatggatcc tccatggatcc tccatggatcc tccatggatcc
2581 ctacgcaacta acttacgt tggaaatcgaa acacccaaac aacatcgat ttttttttttgcgtt
2641 aagcgaatgc agatggaga aacccctccga gactggatcc tccatggatcc tccatggatcc
2701 ctacgcctata gcccgaacgag cagctcagag tgggttgcgtt ttttttttttgcgtt
2761 atgcttcaac taacatgttgc gcttgcggc gttcatgtc tccatggatcc tccatggatcc
2821 cacttccagg ctaacacagt cagaatcgaa aacgtactct caacaggctcg ttttttttttgcgtt
2881 gaagtttgc ggcattctgg ctacacaata aacatcgat ttttttttttgcgtt
2941 ctactacgtc aaaattttt cccatgttgc gttcatgtc tccatggatcc tccatggatcc
3001 tcttagagcga tccgggtatct cggggaaaagg gttcatgtc tccatggatcc tccatggatcc
3061 ctttaaaaat aaaaacataat tactcgtgc ttttttttttgcgtt
3121 tgcttacatc acaacaaaaaa ctgtttaac aacatcgat ttttttttttgcgtt
3181 tgaacattat ctgttataa ttattgtataa aacatcgat ttttttttttgcgtt
3241 gatgcctatc attgggttgc atgaacttgc aacatcgat ttttttttttgcgtt
3301 aggtaaacgc cattgtcagc aaattgtatcc aacatcgat ttttttttttgcgtt
3361 aatgtttaatt ctgcgtgacc ctgagactgt aacatcgat ttttttttttgcgtt
3421 attaaggtaa ggtggataca catcttgcgtca aacatcgat ttttttttttgcgtt
3481 cattaggcac cccaggctt acactttatgc aacatcgat ttttttttttgcgtt
3541 agcggataac aatttcacac agggaaacgc aacatcgat ttttttttttgcgtt
3601 aacccctactt aaaaacggaaaca aacatcgat ttttttttttgcgtt
3661 agtggatccc cccggctgca ggaatttgcgt aacatcgat ttttttttttgcgtt
3721 gggggcccg gtacccaaat cggccctatag aacatcgat ttttttttttgcgtt
3781 cgtttacaa cgtcggtact gggaaaaccc aacatcgat ttttttttttgcgtt
3841 acatcccccatt ttcggccagct ggggtatag aacatcgat ttttttttttgcgtt
3901 acagtggcgc agcgtatgc gacgtatggaa aacatcgat ttttttttttgcgtt
3961 gctgttaatt ttgttaat cgtcttgcgtca aacatcgat ttttttttttgcgtt
4021 ccttataaaat caaaagaata gaccggatata aacatcgat ttttttttttgcgtt
4081 agtccactat taaaagaaacgt ggcactccaaat aacatcgat ttttttttttgcgtt
4141 gatggcccac tactccgggta tcatatgaca aacatcgat ttttttttttgcgtt

4201 ttacccaaaat cattagggga ttcatcgtg
 4261 aggaaagctt atgatgtga tggcttaaa
 4321 caatacatgc gaaaaaccta aaagagcttg
 4381 accgcggctt ttatttgc ttgaaagata
 4441 aatcttctt atcgtaaaaatggccctt
 4501 gaataacatc atttggac gaaataacta
 4561 ttgggggtt aacatgaagg toatcgatag
 4621 aataatccaa atccagccat cccaaattgg
 4681 taatggcca ataacaccgg ttgcattgtt
 5 4741 ttgctgatga ctcttgggat gatagacat
 4801 accaccagcc aataaaaatta aacaggaa
 4861 aaaggcaaat gcatactat ctgcataaa
 4921 ttatgtgcta ttcttcgtc cacaaggct
 4981 gtaatgaaaa gccaaccatc atgttacatc
 5041 gtgctggatt ggctatcaat ggcgtgaaa
 5101 ttagtgtatac cgatcagctt ttgttccctt
 5161 tcatggtcat agctttcc tggtaaaaat
 5221 cgagccggaa gcataaaagtg taagcctgg
 5281 attgcgttgc gtcactgtcc cgcttccag
 5341 ttaatgcggcc aacgcgcggg gagaggcgg
 5401 ctcactgact cgctgcgtc ggctgttgc
 5461 gggtaatac gtttatccac agaatcagg
 5521 gcccagcaaa aggccaggaa cgtaaaaaag
 5581 cgccccctg acgagcatca caaaaatcga
 5641 ggactataaa gataccaggc gtttccccc
 5701 accctgcgcg ttagccgata cctgtccgg
 5761 catagtcac gtttaggtt ttcgttgc
 5821 gtgcacaaac cccctgttca gcccggacc
 5881 tccaaacccgg taagacacga ttatcgcca
 5941 agagcgaggt atgtaggccg tgctacagag
 6001 actagaagga cagtatttgg tatctgcgt
 6061 gtttgtagct cttgtatccgg caaacaaacc
 6121 aagcagcaga ttacgcgcag aaaaaaagga
 6181 gggtgtacg ctcagggaa cggaaaactca
 6241 aaaaggatct tcacatgtat cttttttat
 6301 atatatgtt aacattggc tgacagttac
 6361 gggatctgtc tatttgcgtt atccatagt
 6421 atacgggagg gtttaccatc tggccccagt
 6481 cccggccctcag atttatcagc aataaaccag
 6541 cctgcacactt tateccgcctc catccagtt
 6601 agttcgccag ttaatagttt ggcacaaacgtt
 6661 cgtcgctgt ttgtatggc ttatccagc
 6721 ttagccccca tgggttgcaaa aaaaagggtt
 6781 agtaagtggg cccagtggtt atcaactcat
 6841 gtcatgccccat cccgtaaatg cttttctgt
 6901 gaatagtgtt tgccggcacc gatgtgtct
 6961 ccacatagca gaatctttaaa agtgctcatt
 7021 tcaaggatct taccgcgtt gatccag
 7081 ttctcagcat cttttacttt caccacgg
 7141 gccgcggaaaa agggataaag ggcacacgg
 7201 caatattattt gaaadattta tcagggttat
 7261 atttagaaaa ataaacaaat aggggttccg

55 SEQ ID NO:57 (pTnMod)

60	CTGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGGGGGTGT	GGTGGTTACG	50
	CGCAGCGTGA	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CTCCCTTCGC	100
	TTTCTTCCCT	TCCTTTCTCG	CCACGTTCGC	CGGCATCAGA	TTGGCTATTG	150
	GCCATTGCAT	ACGTTGTATC	CATATCATAA	TATGTACATT	TATATTGGCT	200
	CATGTCCAAC	ATTACCGCCA	TGTTGACATT	GATTATTGAC	TAGTTATTAA	250
	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT	AGCCCATATA	TGGAGTTCCG	300
	CGTTACATAA	CTTACGGTAA	ATGGCCCGCC	TGGCTGACCG	CCCAACGACC	350
	CCCGCCCCATT	GACGTCAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	400
65	GGGACTTTCC	ATTGACGTCA	ATGGGTGGAG	TATTTACGGT	AAACTGCCA	450
	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTACGCC	CCTATTGACG	500

TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA CATGACCTTA 550
 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600
 CATGGTGATG CGGTTTGGC AGTACATCAA TGGCGTGGA TAGCGGTTG 650
 ACTCACGGGG ATTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTG 700
 5 TTTGGCACC AAAATCAACG GGACTTCCA AAATGTCGA ACAACTCCGC 750
 CCCATTGACG CAAATGGGGC GTAGGCGTGT ACGGTGGGAG GTCTATATAA 800
 GCAGAGCTCG TTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850
 TGTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900
 GGAACGGTGC ATTGGAACGC GGATTCCCCG TGCCAAGAGT GACGTAAGTA 950
 10 CCGCCTATAG ACTCTATAGG CACACCCCTT TGGCTTAT GCATGCTATA 1000
 CTGTTTTGG CTGGGGCCT ATACACCCCA GCTTCCCTAT GCTATAGGTG 1050
 ATGGTATAGC TTAGCCTATA GGTGTGGTT ATTGACCATT ATTGACCACT 1100
 CCCCTATTGG TGACGATACT TTCCATTACT AAATCCATAAC ATGGCTCTT 1150
 GCCACAACTA TCTCTATTGG CTATATGCCA ATACCTGTC CTTCAAGAGAC 1200
 15 TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCCATT ATTATTTACA 1250
 AATTACACATA TACAACAACG CCGTCCCCCG TGCCCCGAGT TTTTATTAAA 1300
 CATAGCGTGG GATCTCCACG CGAACATCTCGG GTACGTGTTG CGGACATGGG 1350
 CTCTTCTCCG GTAGGGGGGG AGCTTCCACA TCCGAGCCCT GGTCCTCATGC 1400
 CTCCAGCGGC TCATGGTCGC TCGGCAGCTC CTTGCTCCTA ACAGTGGAGG 1450
 20 CCAGACTTAG GCACAGCACA ATGCCACCA CCACCAAGTGT GCCGCACAAG 1500
 GCGTGGCGG TAGGGTATGT GTCTGAAAAT GAGCGTGGAG ATTGGGCTCG 1550
 CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600
 GCAGCTGAGT TGTGTGATTIC TGATAACAGT CAGAGGTAAC TCCCCTTGCG 1650
 GTGCTGTTAA CGGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700
 25 CGCGCGCGCC ACCAGACATA ATAGCTGACA GACTAACAGA CTGTTCCCTT 1750
 CCATGGGTCT TTCTGCGAT CACCGTCGGA CCATGTGTGA ACTTGATATT 1800
 TTACATGATT CTCTTACCA ATTCTGCCCG GAATTACACT TAAAACGACT 1850
 CAACAGCTTA ACGTTGGCTT GCCACGCCATT ACTTGACTGT AAAACTCTCA 1900
 CTCTTACCGA ACTTGGCCGT AACCTGCCAA CCAAAGCAG AACAAAACAT 1950
 30 AACATCAAC GAATCGACCG ATTGTTAGGT AATCGTCACC TCCACAAAGA 2000
 GCGACTCGCT GTATACCGTT GGCATGCTAG CTTTATCTGT TCGGGAATAC 2050
 GATGCCATT GTACTTGTG ACTGGCTCTGA TATTGCGTAG CAAAAACGAC 2100
 TTATGGTATT GCGAGCTTCA GTCGCACTAC ACGGTCGTTG TGTTACTCTT 2150
 TATGAGAAAG CGTTCCCGCT TTCAGAGCAA TGTCAAAGA AAGCTCATGA 2200
 35 CCAATTCTA GCGGACCTTG CGAGCATTCT ACCGAGTAAC ACCACACCGC 2250
 TCATTGTCAG TGATGCTGGC TTTAAAGTGC CATGGTATAA ATCCGTTGAG 2300
 AAGCTGGTT GGTACTGGTT AAGTCGAGTA AGAGGAAAAG TACAATATGC 2350
 AGACCTAGGA GCGGAAAACT GGAAACCTAT CAGCAACTTA CATGATATGT 2400
 CATCTAGTCA CTCAAAGACT TTAGGCTATA AGAGGCTGAC TAAAAGCAAT 2450
 40 CCAATCTCAT GCGAAATTCT ATTGTTAAA TCTCGCTCTA AAGGCCGAAA 2500
 AAATCAGCGC TCGACACCGA CTCATTGTC CCAACCGCTA CCTAAAATCT 2550
 ACTCAGCGTC GGCAAAAGGAG CCATGGGTTC TAGCAACTAA CTTACCTGTT 2600
 GAAATTGCAA CACCCAAACA ACTTGTTAAT ATCTATTGCA AGCGAATGCA 2650
 GATTGAAGAA ACCTTCCGAG ACTTGAAAAG TCCCTGCTAC GGACTAGGCC 2700
 45 TACGCCATAG CGGAACGAGC AGCTCAGAGC GTTTTGATAT CATGCTGCTA 2750
 ATCGCCCTGA TGCTTCAACT AACATGTTGG CTTGCGGGCG ITTCATGCTCA 2800
 GAAACAAGGT TCGGACAAGC ACTTCCAGGC TAACACAGTC AGAAATCGAA 2850
 ACGTACTCTC AACAGTTCCC TTAGGCATGG AAGTTTGCG GCATTCTGGC 2900
 TACACAATAA CAAGGGAAAGA CTTACTCGTG GCTGCAACCC TACTAGCTCA 2950
 50 AAATTATTTC ACACATGGTT ACGCTTGGG GAAATTATGA TAATGATCCA 3000
 GATCACTTCT GGCTAAATAA AGATCAGAGC TCTAGAGATC TGTGTGTTGG 3050
 TTTTTGTGG ATCTGCTGTG CCTTCTAGTT GCCAGCCATC TGTGTGTTGC 3100
 CCCTCCCCCG TGCTTCTCTT GACCCTGGAA GGTGCCACTC CCACTGTCT 3150
 TTCCTAATAA ATGAGGAAA TTGCATCGCA TTGCTGTAGT AGGTGTCATT 3200
 55 CTATTCTGGG GGGTGGGGTG GGGCAGCACA GCAAGGGGA GGATTGGGAA 3250
 GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG GTACCTCTCT 3300
 CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCGGTAC CTCTCTCTCT 3350
 CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CGGTACCAAGG TGCTGAAGAA 3400
 TTGACCCGGT GACCAAAGGT GCCTTTATC ATCACTTTAA AAATAAAAAA 3450
 60 CAATTACTCA GTGCCGTGTTA TAAGCAGCAA TTAATTATGA TTGATGCCCTA 3500
 CATCACAAACA AAAACTGATT TAACAAATGG TTGGTCTGCC TTAGAAAGTA 3550

TATTTGAACA TTATCTTGAT TATATTATTG ATAATAATAA AAACCTTATC 3600
 CCTATCCAAG AAGTGTGATGCC TATCATGGT TGGATGAAAC TTGAAAAAAA 3650
 TTAGCCTTGA ATACATTACT GGTAAGGTAA ACGCCATTGT CAGCAAATTG 3700
 ATCCAAGAGA ACCAACTTAA AGCTTCTCG ACGGAATGTT AATTCTCGTT 3750
 5 GACCCCTGAGC ACTGATGAAT CCCCTAATGA TTTGGTAAA AATCATTAAG 3800
 TTAAGGTGGA TACACATCTT GTCATATGAT CCCGGTAATG TGAGTTAGCT 3850
 CACTCATTAG GCACCCCGAGG CTTTACACTT TATGCTTCCG GCTCGTATGT 3900
 TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC 3950
 CATGATTACG CCAAGCGCGC AATTAACCC CACTAAAGGG ACACAAAGCT 4000
 10 GGAGCTCCAC CGCGGTGGCG GCGCCTCTAG AACTAGTGGA TCCCCCGGGC 4050
 TGCAGGAATT CGATATCAAG CTTATCGATA CCGCTGACCT CGAGGGGGGG 4100
 CCCGGTACCC AATTGCGCCCT ATAGTGTGTC GTATTACGCG CGCTCACTGG 4150
 CCGTCGTTTT ACAACGTGCGT GACTGGAAA ACCCTGGCGT TACCCAACCTT 4200
 AATCGCCTTG CAGCACATCC CCCCTTCGCG AGCTGGCGTA ATAGCGAAGA 4250
 15 GGCCCGCACC GATCGCCCTT CCCAACAGTT GCGCAGCCTG ATGGCGAAT 4300
 GGAAATTGCTA AGCGTAAATA TTTGTTAAA ATTGCGCTTA AATTTTTGTT 4350
 AAATCAGCTC ATTTTTTAAC CAATAGGCCG AAATCGGCAA AATCCCTTAT 4400
 AAATCAGGAAAG AATAGACCGA GATAGGGTT AGTGTGTT CAGTTTGGAA 4450
 CAAGAGTCCA CTATTAAGA ACGTGGACTC CAACGTCAA GGGCGAAAAA 4500
 20 CCGTCTATCA GGGCGATGGC CCACTACTCC GGATCATAT GACAAGATGT 4550
 GTATCCACCT TAACTTAATG ATTTTTACCA AAATCATTAG GGGATTCACTC 4600
 AGTGTCTCAGG GTCAACGAGA ATTAACATTC CGTCAGGAAA GCTTATGATG 4650
 ATGATGTGCT TAAAAACTTA CTCATGGCT GGTTATGCAT ATCGCAATAC 4700
 ATGCGAAAAAA CCTAAAAGAG CTTGCCGATA AAAAAGGCCA ATTTATTGCT 4750
 25 ATTTACCGCG GCTTTTATT GAGCTGAAA GATAAATAAA ATAGATAGGT 4800
 TTTATTGAA GCTAAATCTT CTTTATCGTA AAAAATGCC TCTTGGGTTA 4850
 TCAAGAGGGT CATTATATT CGCGGAATAA CATCATTG TGACGAAATA 4900
 ACTAAGCACT TGTCTCCCTGT TTACTCCCT GAGCTTGAGG GGTTAACATG 4950
 AAGGTATCG ATAGCAGGAT AATAATACAG TAAAACGCTA AACCAATAAT 5000
 30 CCAAATCCAG CCATCCAAA TTGGTAGTGA ATGATTATAA ATAACAGCAA 5050
 ACAGTAATGG GCCAATAACA CCGGTTGCAT TGGTAAGGCT CACCAATAAT 5100
 CCCTGTAAAG CACCTTGCTG ATGACTCTTT GTTTGGATAG ACATCACTCC 5150
 CTGTAATGCA GGTAAAGCGA TCCCCACCACG AGCCAATAAA ATTAAAACAG 5200
 GGAAAACCTAA CCAACCTTCA GATATAAACG CTAAAAGGC AAATGCACTA 5250
 35 CTATCTGCAA TAATCCGAG CAGTACTGCC GTTTTTTCGC CCATTTAGTG 5300
 GCTATCTTC CTGCCACAAA GGCTTGGAT ACTGAGTGT AAAGACCAAG 5350
 ACCCGTAATG AAAAGCCAAC CATCATGCTA TTTCATCATCA CGATTTCTGT 5400
 AATAGCACCA CACCGTGTG GATTGGCTAT CAATGCGCTG AAATAATAAT 5450
 CAACAAATGG CATGTTAAA TAAGTGTGTT ATACCGATCA GCTTTTGTTC 5500
 40 CCTTTAGTGA GGTTAAATT CGCGCTTGGC GTAATCATGG TCATAGCTGT 5550
 TTCCCTGTGT AAATGTTAT CCGCTCACAA TTCCACACAA CATACTGAGCC 5600
 GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC 5650
 ATTAATTGCG TTGCGCTCAC TGCCCGCTT CCAGTCGGGA AACCTGTGCGT 5700
 GCCAGCTGCA TTAATGAATC GGCAACGCG CGGGGAGAGG CGGTTTGCGT 5750
 45 ATTGGGCCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTGCGT 5800
 TCGGCTGCGG CGACCGGTAT CAGCTCACTC AAAGGGCGGTATACGGTTAT 5850
 CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG 5900
 CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT TTTTCCATAG 5950
 GCTCCGCCCC CCTGACGAGC ATCACAACAA TCGACGCTCA AGTCAGAGGT 6000
 50 GGCAGAAACCC GACAGGACTA TAAAGATACC AGGCCTTTC CCCTGGAAAGC 6050
 TCCCTCGTGC GCTCTCCGT TCCGACCCCTG CCGCTTACCG GATACTGTC 6100
 CGCCTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA 6150
 GGTATCTCAG TTGCGTGTAG GTGCGTTCGCT CCAAGCTGGG CTGTGTGAC 6200
 GAACCCCCCG TTCAGCCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT 6250
 55 TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG 6300
 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG 6350
 AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG 6400
 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT 6450
 CCGGCAAACCA ACCAACCGCT GGTAGCGGTG GTTTTTTGT TTGCAAGCAG 6500
 60 CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCCT TGATCTTITC 6550
 TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTGG 6600

TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTT AAATTAAGAA 6650
 TGAAGTTTA AATCAATCTA AAGTATATAA GAGTAAACCTT GGTCTGACAG 6700
 TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTC 6750
 GTTCATCCAT AGTTGCCCTGA CTCCCCGTCG TGTAGATAAC TACGATAACGG 6800
 5 GAGGGCTTAC CATCTGGCCC CAGTGTGCA ATGATACCGC GAGACCCACG 6850
 CTCACCCGGCT CCAGATTAT CAGCAATAAA CCAGGCCAGCC GGAAGGGCCG 6900
 AGCGCAGAAG TGGTCCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAAT 6950
 TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTGCCTGCAA 7000
 CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTGGTA 7050
 10 TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGGCAGT TACATGATCC 7100
 CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCCGGTCTC CGATCGTTGT 7150
 CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC 7200
 ATAATTCTCT TACTGTCAATG CCATCCGTA GATGCTTTTC TGTGACTGGT 7250
 GAGTACTCAA CCAAGTCATT CTGAGAAATAG TGTATGCGGC GACCGAGTTG 7300
 15 CTCTTGCCCG GCGTCAATAC GGGATAATAC CGGCCACAT AGCAGAACTT 7350
 TAAAAGTGCT CATCAATTGGA AAACGTTCTT CGGGCGAAA ACTCTCAAGG 7400
 ATCTTACCGC TGTTGAGATC CAGTTCCATG TAACCCACTC GTGCACCCAA 7450
 CTGATCTTCA GCATCTTTA CTTTACCCAG CGTTCTGGG TGAGCAAAA 7500
 20 CAGGAAGGCA AAATGCCCA AAAAGGGAA TAAGGGCGAC ACGGAAATGT 7550
 TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG 7600
 TTATTGTCTC ATGAGCGGAT ACATATTGAA ATGTTTTAG AAAAATAAAC 7650
 AAATAGGGGT TCCGCCACAA TTTCCCCGAA AAGTGCCAC 7689

SEQ ID NO:58 (a Kozak sequence)

25 ACCATGG

SEQ ID NO:59 (a Kozak sequence)

ACCATGT

30 SEQ ID NO:60 (a Kozak sequence)

AAGATGT

SEQ ID NO:61 (a Kozak sequence)

ACGATGA

35 SEQ ID NO:62 (a Kozak sequence)

AAGATGG

SEQ ID NO:63 (a Kozak sequence)

40 GACATGA

SEQ ID NO:64 (a Kozak sequence)

ACCATGA

45 SEQ ID NO:65 (a Kozak sequence)

ACCATGT

SEQ ID NO:66 (conalbumin polyA)

50 tctgccattg ctgcttcctc tgcccttctt cgtcaactctg aatgtggctt cttcgctact
 gcccacagcaa gaaataaaaat ctcaacatct aatgggttt cctgagggtt ttcaagagtc
 .gttaacgcaca ttccttcccc agcaccctt gctgcaggcc agtgcacggc accaacttgg
 ctactgctgc ccatgagaga aatccagttc aatattttcc aaagcaaaat ggattacata
 tgccttagat cctgatataac aggcgtttgt attatctagt gctttcgctt cacccagatt
 atcccatattgc ctccc

55 SEQ ID NO:67 (synthetic polyA)

GGCGCCCTGGATCCAGATCACTTCTGGCTAATAAAAGATCAGAGCTCTAGAGATCTGTGTGGTTTTT
 TGTGGATCTGCTGTGCCCTCTAGTTGCCAGCCATCTGTTGCCCTCCCCGTGCCTCCTGACC

5 SEQ ID NO:68 (avian optimized polyA)

10 ggggatcgctcttagagcgtccgggatctcggggaaaagcggttggtgaccaaaggtgccttttatcatcactttaaaaataaaaaacaattaactcagtgcctgttataagcagcaattattatgattgatgcctacatcacaacaaaaatcgatttaacaaatggttggtctgccttagaaagtatatttgaacattatcttgattatatttgataataaaaaaccttacccctatccaagaagtatgcctatcattgggttggatgaacttgaaaaaaatttagccttgaatacattactggtaaqgtaaacgcctattgtcagcaattgtatccaaqgaaaccaaa

15 SEO ID NO:69 (vitellogenin promoter)

20	TGAATGTGTT	CTTGTGTAT	CAATATAAAT	CACAGTTAGT	GATGAAGITG	GCTGCAAGCC
	TGCATCAGTT	CAGCTACTTG	GCTGCATTTT	GTATTTGGTT	CTGTAGGAAA	TGCAAAAGGT
	TCTAGGCTGA	CTTGCACCTTC	TATCCCTCTT	GCCTTACTGC	TGAGAACTCTC	TGCAAGGTTTT
	AATTGTTCAC	ATTTTGCTCC	CATTACTTT	GGAAGATAAA	ATATTTACAG	AATGCTTATG
	AAACCTTTGT	TCATTTAAAA	ATATTCCCTGG	TCAGCGTGCAC	CGGAGCTGAA	AGAACACATT
	GATCCCCTGTA	TTTCAATAAA	TACATATGTT	CCATATATTG	TTTCTCAGTA	GCCTCTTAAA
25	TCATGTGCGT	TGGTGCACAT	ATGAATAACAT	GAATAGCAA	GGTTTATCTG	GATTACGCTC
	TGGCCTGCAG	GAATGGCCAT	AAACCAAAGC	TGAGGGAAAGA	GGGAGAGTAT	AGTCAATGTA
	GATTATACTG	ATTGCTGATT	GGGTATTATAT	CAGCTAGATA	ACAACATTGGG	TCAGGTGCCA
	GGTCAACATA	ACCTGGGCAA	AACCAGTCTC	ATCTGTGGCA	GGACCATGTA	CCAGCAGGCCA
	GCCGTGACCC	AATCTAGGAA	AGCAAGTAGC	ACATCAATT	TAAATTATT	GTAATGCGCG
30	TAGTAGAAAGT	GTTTGTGATT	GATACATTGA	AACTTCTGGT	CAATCAGAAA	AAGGTTTTTT
	ATCAGAGATG	CCAAGGTATT	ATTGATTTT	CTTATTCTCG	CGTGAAGAGA	ATTATGATT
	GCAAAAAGAG	GAGTGTAC	ATAAAATGAT	AAAAAAACTTG	AGGAATTCA	CAGAAAACAG
	CCACGTGTTC	CTGAACATTC	TTCCATAAAA	GTCTCACCAT	GCCTGGCAGA	GCCCTATTCA
	CCTTCGCT					

35 SEO ID NO:70 (fragment of ovalbumin promoter - chicken)

40	GAGGTCTAGAAT GGTTTCTTTA CTGTTTGTC ATTCTATTAT TTCAATAACAG AACAAATAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTCACAA TTCCCTGTC ATCTGCCAGG CCATTAAGTT ATTCTATGGAA GATCTTGAG GAACACTGCA AGTTCATATC ATAAACACAT TTGAAATTGA GTATTGTTT GCATTGTATG GAGCTATGTT TTGCTGTATC CTCAGAAAAA AAGTTTGTAA TAAAGCATTG ACACCCATAA AAAGATAGAT TTAAATATTC CAGCTATAGG AAAGAAAGTG CGTCTGCTCT TCACTCTAGT CTCAGTTGGC TTCTTCACAT GCATGCTTCT
45	TTATTTCTCC TATTTTGTCA AGAAAATAAT AGGTACACGTC TTGTTCTCAC TTATGTCTG CCTAGCATGG CTCAGATGCA CGTTGTAGAT ACAAGAAGGA TCAAATGAAA CAGACTTCTG GTCTGTACT ACAACCATAG TAATAAGCAC ACTAACTAAT AATTGCTAAT TATGTTTCC ATCTCTAAGG TTCCCACATT TTTCTGTTT CTTAAAGATC CCATTATCTG GTTGTAACTG AAGCTCAATG GAACATGAGC AATATTCTCC AGTCTTCTCT CCCATCCAAC AGTCCTGATG GATTAGCAGA ACAGGCAGAA AACACATTGT TACCCAGAAAT TAAAAGCTAA TATTTGCTCT CCATTCAATC CAAAATGGAC CTATTGAAAC TAAAATCTAA CCCAATCCCA TTAAATGATT TCTATGGCGT CAAAGGTCAA ACTTCTGAAG GGAACCTGTC GGTGGGTCAAC AATTCAAGGCT ATATATTCCC CAGGGCTCAG
50	

SEQ ID NO:71 (chicken ovalbumin enhancer)
ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag
cttgacctga tacctgatt tcttcaaact ggggaaacaa cacaatccca caaaacagct
cagagagaaa ccatcaactga tggctacagc accaaggat gcaatggcaa tccattcgcac
attcatctgt gacctgagca aaatgttta tctctccatg aatgttgtct tctttccctc
60

atgaaaaggc aatttccaca ctcacaatat gcaacaaaaga caaacagaga acaattaatg
 tgctccctcc taatgtcaaa attgtatggg caaagaggag aacaaaatct caagttctga
 gtaggttttta gtgattggat aagaggctt gacctgtgag ctcacctgga cttcatatcc
 ttttggataa aaagtcttt tataacttcc aggtctccga gtctttatcc atgagactgt
 5 tggtttaggg acagacccac aatgaaatgc ctggcatagg aaagggcagg agagccttag
 ctgaccccttt ctgggacaa gcattgtcaa acaatgtgtg acaaaaactat ttgtactgct
 ttgcacagct gtgctggca gggcaatcca ttgcaccta tcccaggtaa ctttccaact
 gcaagaagat tggcttac tctctctaga

10 SEQ ID NO:72 (5' untranslated region)
 GTGGATCAACATAACAGCTAGAAAGCTGTATTGCCTTAGCACTCAAGCTCAAAAGACAACTCAGAGTTCA
 ACC

15 SEQ ID NO:73 (putative cap site)
 ACATACAGCTAG AAAGCTGTAT TGCCTTAGC ACTCAAGCTC AAAAGACAAAC TCAGAGTTCA

20 SEQ ID NO:74 (Chicken Ovalbumin Signal Sequence)
 ATG GGCTCCATCG GCGCAGCAAG CATGGAATT TGTTTGATG TATTCAAGGA GCTCAAAGTC
 CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TCATGTCAAGC TCTAGCCATG
 GTATACCTGG GTGCAAAAGA CAGCACCAGG ACACAGATAA ATAAGGTTGT TCGCTTTGAT
 25 AAACTTCCAG GATTGGAGA CAGTATTGAA GCTCAGTGTG GCACATCTGT AAACGTTCAC
 TCTTCACCTA GAGACATCCT CAACCAAATC ACCAAACCAA ATGATGTTA TTGTTTCAGC
 CTTGCCAGTA GACTTTATGC TGAAGAGAGA TACCCAATCC TGCCAGAATA CTTGCAGTGT
 GTGAAGGAAC TGTATAGAGG AGGCTTGAA CCTATCAACT TCACAAACAGC TGCAGATCAA
 GCCAGAGAGC TCATCAATTG CTGGGTAGAA AGTCAGACAA ATGGAATTAT CAGAAATGTC
 CTTCAGCCAA GCTCCGTGGA TTCTCAAAC GCAATGGTTC TGGTTAATGC CATTGTTTC
 30 AAAGGACTGT GGGAGAAAAC ATTTAAGGAT GAAGACACAC AAGCAATGCC TTTCAGAGTG
 ACTGAGCAAG AAAGCAAACG TGTGCAGATG ATGTAACAGA TTGGTTTATT TAGAGTGGCA
 TCAATGGCTT CTGAGAAAAT GAAGATCTG GAGCTTCCAT TTGCCAGTGG GACAATGAGC
 ATGTTGGTGC TGTGCTGCA TGAAGTCCTCA GGCCTTGAGC AGCTTGAGAG TATAATCAAC
 TTTGAAAAAC TGACTGAATG GACCACTCT AATGTTATGG AAGAGAGGAA GATCAAAGTG
 35 TACTTACCTC GCATGAAGAT GGAGGAAAAA TACAACCTCA CATCTGTCTT AATGGCTATG
 GGCATTACTG ACCTGTTTAG CTCTTCAGCC AATCTGTCTG GCATCTCCCTC AGCAGAGAGC
 CTGAAGATAT CTCAGCTGT CCATGCAGCA CATGAGAAA TCAATGAAGC AGGCAGAGAG
 GTGGTAGGGT CAGCAGAGGC TGGAGTGGAT GCTGCAAGCG TCTCTGAAGA ATTTAGGGCT
 GACCATCCAT TCCCTCTCTG TATCAAGCAC ATCGCAACCA ACGCCGTTCT CTTCTTGCG
 40 AGATGTGTTT CCCCT

SEQ ID NO:75 (Chicken Ovalbumin Signal Sequence-shortened approx.
 50bp)

45 ATG GGCTCCATCG GCGCAGCAAG CATGGAATT TGTTTGATG TATTCAAG

SEQ ID NO:76 (Chicken Ovalbumin Signal Sequence-shortened approx.
 100bp)

50 ATG GGCTCCATCG GCGCAGCAAG CATGGAATT TGTTTGATG TATTCAAGGA GCTCAAAGTC
 CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TC

SEQ ID NO:77 (vitellogenin targeting sequence)

55 ATGAGGGGGATCATACTGGCATTAGTGCTACCCCTGTAGGCAGCAGAAGTTGACATTGGT

SEQ ID NO:78 (p146 protein)
 KYKKALKKLAKLL

60 SEQ ID NO:79 (p146 coding sequence)

AAATACAAAAAAGCACTGAAAAAACTGGCAAAACTGCTG

5 SEQ ID NO:80 (pro-insulin sequence)
TTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGAAGCTCTACCTAGTGTGCGGGAAACGAGGC
TTCTTCTACACACCAAGACCCGCGGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGCGGG
GGCCCTGGTGCAGGCAGCCTGCAGCCCTGGCCCTGGAGGGTCCCTGCAGAAGCGTGGCATTGTGGAA
CAATGCTGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTCTGCAACTAG

10 SEQ ID NO:81 (spacer)
(GPGG)_x

15 SEQ ID NO:82 (spacer)
GPGGGPGGGPGG

20 SEQ ID NO:83 (spacer)
GGGGSGGGGSGGGGS

25 SEQ ID NO:84 (spacer)
GGGGSGGGGSGGGGSGGGGS

30 SEQ ID NO:85 (repeat domain in TAG spacer sequence)
Pro Ala Asp Asp Ala

35 SEQ ID NO:86 (TAG spacer sequence)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp

40 SEQ ID NO:87 (gp41 epitope)
Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu

45 SEQ ID NO:88 (polynucleotide sequence encoding gp41 epitope)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly
Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Lys

50 SEQ ID NO:89 (enterokinase cleavage site)
DDDDDK

55 SEQ ID NO:90 (TAG sequence)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys
Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys

55 SEQ ID NO:91 (altered transposase Hef forward primer)
ATCTCGAGACCATGTGTGAACCTTGATATTTACATGATTCTCTTAC

SEQ ID NO:92 (altered transposase Her reverse primer)
GATTGATCATTATCATAATTCCCCAAAGCGTAACC

5 SEQ ID NO:93 (Xho I restriction site)
CTCGAG

10 SEQ ID NO:94 (Bcl I restriction site)
TGATCA

SEQ ID NO:95. (CMVf-NgoM IV primer)
TTGCCGGCATCAGATTGGCTAT

15

SEQ ID NO: 96 (Syn-polyAr-BstE II primer)
AGAGGTCAACGGGTCAATTCTCAGCACCTGGTA

20

SEQ ID NO:97 (pTnMOD (CMV-prepro-HCPro-Lys-CPA)

25	1	ctgacgcgcc	ctgtagcgcc	gcattaaagcg	cgccgggtgt	ggtgtttaacg	cgccagcgta
	61	ccgctact	tgccagcgcc	ctagcccccg	ctcccttcgc	ttttttccct	tcctttctcg
	121	ccacgttgc	cgccatcaga	ttggctattg	gcatttgcatt	acgttgtatc	catacataaa
	181	tatgttacatt	tatattggct	catgtccaaac	attacccgca	tgttgcatt	gattattgac
	241	tagttattaa	tagtaatcaa	ttacggggct	attatgttcat	agccccatata	tggagttcog
30	301	cgttacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	ccggccccc
	361	gacgtcaata	atgacgtatg	ttccccatagt	aacgcataa	gggactttcc	attgacgtca
	421	atgggtggag	tatttacggt	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc
	481	aagtacgccc	cctattgacg	tcaatgacgg	taatggcccc	gcctggcatt	atgcccagta
	541	catgacccta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgcatttac
	601	catggtgatg	cgggtttggc	agtacatcaa	tgggctgga	tagcggtttg	actcacgggg
	661	atttccaagt	ctccacccca	ttgacgtcaa	tggaggtttg	ttttggcacc	aaaatcaacg
35	721	gagacttcca	aaatgtcgta	aaacactccgc	ccccattgacg	caaatggccg	gtaggcgtgt
	781	acgggtggag	gtctatataa	gcagagctcg	ttagtgaac	cgtcagatcg	cctggagacg
	841	ccatccacgc	tgttttggacc	tccatagaag	acacccggac	cgatccagcc	tcggggcccg
	901	ggaacgggtc	atttggaaacgc	ggattccccg	tgccaaagagt	gacgtaaagta	ccgcctataag
	961	actctataggg	cacccccc	tggcttttat	gcatgtata	ctgtttttgg	cttggggcct
40	1021	atacaccccc	gttcccttat	gtctataggt	atgttatagc	ttagctata	ggtgtgggtt
	1081	attgaccatt	attgaccact	ccccctattgg	tgacgatact	ttccattact	aatccataac
	1141	atggcttctt	gccacaacta	tctcttattgg	ctatatgcca	atactctgtc	cttcagagac
	1201	tgacacggac	tctgttattt	tacaggatgg	ggttcccattt	atttatttaca	aatttcacata
	1261	tacaacaacg	ccgtcccccg	tgcccgca	tttttattaaa	catagcgtgg	gatctccacg
45	1321	cgaatctcg	gtacgttgc	cgacatggg	cttttctccg	gtacggccgg	agttccacaa
	1381	tccgagccct	ggtcccatgc	ctccagcgcc	tcatgtcg	tcggcagctc	tttgtcccta
	1441	acagtgggg	ccagacttag	gcacagcaca	atgcccacca	ccaccaggatgt	ggccgcacaag
	1501	ggcggtgggg	taggtatgt	gtctgaaaat	gagcggtggag	attgggctcg	cacggctgac
	1561	gagatggaa	gacttaaggc	agcggcagaa	gaagatgcag	cgacgtcgatg	tgttgtatc
50	1621	tgataaagatg	cagggttaac	tccctgtcg	tgctgtttaa	cggtggagggg	cagtgtatc
	1681	tgagcagttac	tcgttgc	cgcgccgccc	accagacata	atagctgaca	gactaaacaga
	1741	ctgttccctt	ccatgggtct	tttctgtcgat	cacccgtcg	ccatgtgt	acttgtatatt
	1801	ttacatgatt	cttttacca	attctgcccc	gaattacact	taaaacgact	caacagctta
	1861	acgttggott	gccaacgatt	acttgcactgt	aaaactctca	ctcttaccga	acttggccgt
55	1921	aacctgccaa	ccaaaggcgag	aacaaaacat	aaatccaaac	gaatcgaccg	attgttaggt
	1981	atacgttca	tccacaaa	gagactcgct	gtataccgtt	ggcatgtcgat	ctttatctgt
	2041	tcgggcaata	cgatggccat	tgtacttgg	gactggctcg	attatcgta	gaaaaaaca
	2101	cttatggtat	tgegagcttc	agtcgtacta	cacggcgctt	ctgttactct	ttatggagaaa
	2161	gggttcccg	tttcagagca	atgttcaaag	aaagtcatg	accaatttct	agccgacc
60	2221	gcgagcattc	taccgagtaa	caccacaccg	cttattgtca	gtgtatgt	ctttaaaagt
	2281	ccatggtata	aatccgttga	gaagctgggt	ttgtacttgg	taagtgcagt	aagaggaaaa
	2341	gtacaaatgt	cagacctagg	agcgaaaaac	tggaaaccta	tcaacactt	acatgtatg
	2401	tcatctatgt	actcaaaagac	ttaggtctat	aagggtctga	ctaaaagcaa	tccaaatctca
	2461	tgccaaattc	tatgttataa	atctcgctt	aaaggccgaa	aaaatcagcg	ctcgacacccg
65	2521	actcattgtc	accacccgtc	acttaaaatc	tacttagcgt	cgccaaaggaa	gcatgggtt
	2581	ctagcaacta	acttacgtt	tgaaattcga	acacccaaac	aacttggtaa	tatctatttc

5	2641	aaggcaatgc	agattgaaga	aaccttcgca	gacttggaaa	gtcctgccta	cggactaggc
	2701	ctacgccata	gccgaacgag	cagctcagag	cgttttgcata	tcatgctgt	aatgcgcctg
	2761	atgcttcaac	taacatgttg	gcttcgggc	gttcatgctc	agaaacaagg	ttgggacaag
	2821	cacttccagg	ctaacacagt	cagaatcga	aacgtactct	caacagtcc	ctttaggcatg
	2881	gaagttttgc	ggcatttcgtt	ctacacaata	acaaggggaa	acttactctg	ggctgcacacc
	2941	ctactagctc	aaaatttttatt	cacacatggt	tacgcctttgg	gaaaattatg	ataatgtatcc
10	3001	agatcacttc	tggctaaataa	aagatcagag	ctctagagat	ctgtgtgtt	gtttttgtg
	3061	gatctgtgt	gccttctagt	tgccagccat	ctgttggtttgc	cccctcccc	gtgccttcct
	3121	tgaccctggta	agggccact	cccactgtcc	tttccataata	aataggggaa	attgcatacg
	3181	attgtctgag	taggtgtcat	tctattttgg	gggggtgggt	ggggcagcac	agcaaggggg
	3241	aggattggga	agacaatagc	aggcactgtgt	gggatgcgg	gggctatgt	ggtagctctc
	3301	tctctctctc	tctctctctc	tctctctc	tctctcggtt	cctctctctc	tctctctctc
	3361	tctctctctc	tctctctctc	teggtaacc	gtgctgaaga	attgaccgg	tgaccaaaaagg
	3421	tgccctttat	catcaacttta	aaaaataaaa	acaattactc	atgcgttgc	ataagcaca
15	3481	attaaattatg	attgtatgc	acatcacaac	aaaaactgtat	ttaacaaatg	gttggctgtc
	3541	cttagaaagt	atatttgaac	attatcttga	ttatattttt	gataataata	aaaacccattat
	3601	ccctatccaa	gaagtgtatc	ctatcattgg	ttggatgaa	cttggaaaaaa	attagccttg
	3661	aatacattac	tggtaaggtt	aacgcatttg	tcagcaattt	gatccaagag	aaccaactta
	3721	aagctttctc	gacggaaatgt	taattttcg	tgaccctgtg	cactgtatgaa	tcccataatg
	3781	attttggtaa	aaatcaattaa	gttaagggtt	ataacatct	tgtcatatgt	tcccgttata
	3841	gtgatgttgc	tcacttcattt	ggcacccca	gtttagatc	ttatgttcc	ggctcgat
	3901	tttggtggaa	tttgtagccg	ataacaattt	cacacaggaa	acagctatga	ccatgattac
	3961	gccaagcgcg	caattaaccc	tcactaaagg	gaacaaaagg	tggagctcca	ccgggggtggc
	4021	ggccgctcta	gaactagtgg	atccccggg	ctgcaggaaat	tcgatatacaa	gttattatcgat
	4081	accgcgtgacc	tcgagcatca	gattggctat	tggccatttg	atacgttgt	tccatatacat
	4141	aatatgtaca	tttatatatttgc	ctcatgttca	acattacggc	catgttgcata	ttgattatttgc
	4201	actagtattt	aatagtaatc	aattacgggg	tcattagttc	atagccatca	tatggagttc
	4261	cgcggttacat	aacttacgg	ttttggccgg	ccttggctgac	cgcccaacga	ccccccccca
	4321	ttgacgtcaa	taatgacgtt	ttgtttccata	gttaacgc	taggactttt	ccattgtacgt
	4381	caatgggtgg	agtatttacg	gttaaactgc	cacttggc	tacatcaagt	gttcatatgt
	4441	tcaagtacgc	ccccatttga	cgtcaatgc	ggtaatggc	ccgcctggca	ttatgc
	4501	tacatgaccc	tatgggactt	tcctacttgg	cagtacatct	acgttattatg	catgcattt
	4561	accatggtga	tgccgttttg	gcagttatc	aatggcgtg	gatagcggtt	tgactc
	4621	ggatttccaa	gttccatcc	cattgcgtc	aatgggat	tgttttggca	ccaaaatcaa
	4681	cgggactttc	caaaatgtcg	taacaactcc	gccccatgt	cgccaaatggg	cggttagcg
	4741	gtacgggtgg	aggctatata	aagcagatgt	cgtttagtga	accgtcatat	cgccgttgc
	4801	cgccatccac	gtcttttgc	ccttcaataga	agacacccgg	gtgacgttgc	taccgcctat
	4861	cgggaaacgg	gtattggaa	ggggattccc	cg	tactgtttt	ggcttggggc
	4921	agactctata	ggcacacccc	tttggctttt	atgcatgtca	gtttagccata	ttatgttgc
	4981	ctatacacc	ccgttccctt	atgcataagg	tgatgttata	cttccatata	taggtgtggg
	5041	ttatttgc	ttatttgc	cttccatattt	gttgcacgata	ctaaatccata	ctaaatccata
	5101	acatggctc	ttgcacacaa	tatcttattt	gttgcata	caatactctg	tccttccatag
	5161	actgacacgg	actctgtat	tttacaggat	gggttccat	ttatttata	caaattccata
	5221	tatataacaa	cgccgtcccc	cgttccgc	gttttttata	aacatagcg	gggatctcca
	5281	cgcgaatctc	gggttacgtgt	tcggacat	ggctttctc	cggttagggc	ggagcttcca
	5341	catccgagcc	ctggccat	gcctccatgc	gttcatgttgc	gctcggcage	tccttgc
	5401	taacagtgaa	ggccagactt	aggcagac	caatgc	caccaccat	gtccgcaca
	5461	aggccgtggc	gtttaggtat	gtgttgc	atagcgtgg	agattgggc	cgacgcgt
	5521	acgcagatgc	aagacttta	gcagccgg	aagaatgtc	aggcagatgt	gttgggtat
	5581	tctgataaga	gtcagatgtt	actccgttgc	cgttgcgtt	aacgggtgg	ggcagtgttag
	5641	tctgacgtt	actctgttgc	ggccggcc	ccacccagaca	taatagctga	cagactaaaca
	5701	gactgttctt	ttccatgggt	cttttgc	gtcacccgt	gttcaatcat	tcatctcg
	5761	acttcttgc	gtgttgcgtt	tacctatata	tctaaat	atatttgc	tattttttt
	5821	taatataattt	cgacgtat	tttctcaagg	atatttttct	tcgttgc	tttgggtct
	5881	gttttgc	cgat	tttgc	ccgaaaggta	cccaatgt	tttgggtct
	5941	tccgggggg	gttggtaaa	ggccgggggg	tcctttagag	cccaatgt	tttgggtct
	6001	tccacttca	gaaacgc	gtatgtgt	gttccgc	cccaatgt	tttgggtct
	6061	ttgggtccgc	gtttaaaa	aaaattgt	ggtgggacaa	cagactatgc	tttgggtct
	6121	aaaggcagat	tcacat	aaagatgt	tcaaaaaaca	cggttatatct	gcaaaatgt
	6181	agcctgaaag	ccgaggacac	agccgtat	tactgtacca	cggggattat	gataacat
	6241	gggggagtt	tcctccccc	gaattgggc	caggaaaccc	ttgttacccgt	tccttcag
	6301	tccaccaagg	ccccatcggt	cttccccc	gcacccct	ccaagagcac	ctctggggc
	6361	acagccggcc	ttgggtcc	gttcaaggac	tacttcc	aaccgggtac	gggtgtcg
	6421	aactcaggcc	ccctgaccc	cggtgtgc	acttcc	cggttccata	gttccat
	6481	ctctacttc	tttagacat	ggtgcacgt	ccctcc	gttggggac	ccagacat
	6541	ctgtcaac	atgtacat	ggccacac	acc	ccaagggaaatgt	tgagcc
	6601	tcttgc	aaaactcac	atgc	ccacccgt	ctgaactct	ggggggaccc
	6661	tcagtcttc	tcttcccc	aaaa	qacaccctca	tgtatctcc	gacccctqag

6721 gtcacatgcg tgggggtgga cgtgagccac gaagaccctg aggtcaagtt caactggtag
 6781 gtggacggcg tggaggtgca taatgccaag acaaagccgc gggaggagca gtacaacagc
 6841 acgtaccgtg tggcagcgt cctcaccgtc ctgaccagg actggctgaa tggcaaggag
 6901 tacaagtgcg aggttccaa caaagccctc ccagecccca tcgagaaaac catctccaa
 5 6961 gccaaagggc agccccgaga accacagggt tacaccctgc ccccatccccg ggatgagctg
 7021 accaagaacc aggtcagct gacgtccctg gtcaaaggt tctatcccag cgacatcgcc
 7081 gtggagtggg agagcaatgg gcggcggag aacaactaca agaccacgc tccegtgtcg
 7141 gactccgacg ctccctctt ctctatcagc aagtcacccg tggacaagag caggtggag
 7201 cagggaaacg tcttctcatg ctccgtatg catgaggatc tgcacaacca ctacacgc
 10 7261 aagagctct cctcttcatg gggtaaaggc ccagagcgc aaaaagtttctatgagctg
 7321 acacagccac cctcgggtgc agtgtccccc ggacaaacgg ccaggatcac ctgctctgga
 7381 gatgcattgc cagaaaaata ttttattgg taccagcaga agtcaggcca ggccctgtg
 7441 gtggatcttct atgaggacag caaacgacc tccggatcc ctgagagatt ctctggctcc
 7501 agctcaggga caatggccac cttgactatc agtggggccc aggtggaaga tgaagggtgac
 15 7561 tactactgtt actcaactga cagcgtgtt tttataggg aggtgtttag cggaggacc
 7621 aagctgaccg ctcttaggtca gcccagggt gccccctcg tcactctgtt cccaccctcc
 7681 tctggaggac ttaaagccaa caaggccaca ctgggtgtc ctataagtga ctcttacccg
 7741 ggaggcgtga cagtggctg gaaggcagat agcagcccg tcaaggcggg agtggagacc
 7801 accacaccct ccaaaacaaag caacaacaag tacggggca gcagctactt gggctgac
 20 7861 cttggcagt ggaagtcaca caaaagctac agtgcacagg tcacgcatga agggagcacc
 7921 gtggagaaga cagtggcccc tgcagaatgt tcaccggga gggagggaa gggccctttt
 7981 gaagggggg gaaacttgcg gccatgactc ctctcggtcc cccgcacgg aacactgtat
 8041 tgcaggggc cctctgcacat tgctgttcc tctgccttc ctcgtcactt tgaatgtggc
 8101 ttcttgcata tccacacag aaaaataaa atctcaacat taaaatgggt ttcttgagat
 25 8161 ttcttgcata tccacacag aaaaataaa atctcaacat taaaatgggt ttcttgagat
 8221 gcaccaactt ggctactgtc gcccattgaga gaaatccagt tcaatatttt ccaaagcaaa
 8281 atggattaca tatgccttag atcctgatta acagggttt ttttattatct gtgtttcgc
 8341 ttccacccaca ttatcccatt gcctccctc gactcgaggg gggggcccg acccaatcc
 8401 ccctatagtg agtcgttta cggcgctca ctggcgctcg ttttacaaacg tctgtactgg
 8461 gaaaaccctg ggttacccca acttaatcgc ctgcacgc atccccctt cggcagctgg
 8521 cgtatagcg aagaggcccg caccgatgcg ctttccaaac agttgcgcag cctgaatggc
 8581 gaatggaaaat ttttgcgtt aatattttt taaaatttcg gttttttttt ttttttttt
 8641 gtcattttt taaccaatag gccgaaatcg gcaaaatccc ttataaatca aaaaataga
 8701 ccgagatagg gttgagtgtt gttccagttt ggaacaagag tccactatta aagaacgtgg
 8761 actccaaatgt caaagggcga aaaaccgtct atcaggggcga tggcccaacta cttccggatc
 8821 atatgadaag atgtgtatcc accttaatct aatgtttttt accaaaatca tttaggggatt
 8881 catcagtgtc cagggtcaac gagaataac attccgtcag gaaagcttat gatgtatgt
 8941 tgctaaaaaa cttaacttgcg ggtctttat gcatatgcga atacatgcga aaaaactaaaa
 9001 aggttgcg gataaaaaaa gccaattttt tgctatttc cgcggctttt tatttgcgtt
 9061 gaaagataaa taaaatagat aggtttttt tgaagctaaa tttttttttt ctgttttttt
 9121 gccccttgg gttatcaaga gggtcattat atttcgcgga ataacatcat ttggtgacga
 9181 aataactaag cacttgcctc ctgtttactc ccctgagctt gaggggttaa catgaagggtc
 9241 atcgatagca ggtataaat acgtaaaaac gctaaaccaaa taatccaaat ccagccatcc
 9301 caaatttggta gtgaatgatt ataataaca gcaaaacgta atggccaaat aacaccgggtt
 9361 gcatgttgc ggttccacaa taatccctgt aaagcacctt gctgtactt ctgtttttgg
 9421 atagacatca ctccctgtta tgcaggtaa gcgatccac caccagccaa taaaataaaa
 9481 acaggaaaaa ctaaccaacc ttccatata aacgtaaaaa aggccaaatgc actactatct
 9541 gcaataaaatc cgagcgtac tggcgcccccc tggccctttt tggccctttt tggccctttt
 9601 caaaggcttg gaataactgat gttttttttt ggttccatcat tttttttttt ttttttttt
 9661 gctatttcatc atcacgtttt ctgtttatgc accacccgtt gctggattttt ctatcaatgc
 9721 gctgtatataa taatcaacaa atggcattttt gttttttttt tttttttttt ttttttttt
 9781 gttcccttta gttgggggtt atttcgcgtt tggcgtatc atggcattttt ttttttttt
 9841 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 9901 aagctggggg tggctaatgc gttggatcactt tttttttttt tttttttttt tttttttttt
 55 9961 ctttccatgc gggaaacccgc tggcgccagg tggccatcat tttttttttt tttttttttt
 10021 gaggggtttt ggttccatcat tttttttttt tttttttttt tttttttttt tttttttttt
 10081 tggcgccgtt gggcgaggcg gttttttttt tttttttttt tttttttttt tttttttttt
 10141 aatcagggggta tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 10201 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60 10261 aaaaatcgacg ctcaaggatcg aggtggcgaa accccgcacgg actataaaaga taccaggcg
 10321 ttcccttgg aagttccctc tttttttttt tttttttttt tttttttttt tttttttttt
 10381 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 10441 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 10501 cccggccgtt gggccatcat tttttttttt tttttttttt tttttttttt tttttttttt
 10561 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 10621 ctacagatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 10681 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 10741 aacaaaccac cgttggatgc ggttccatcat tttttttttt tttttttttt tttttttttt

5	10801	aaaaaggatc	tcaagaagat	cttttgatct	tttctacggg	gtctgacgct	cagtggaaacg
	10861	aaaactcacg	ttaagggatt	ttggtcatga	gattatcaaa	aaggatcttc	acctagatcc
	10921	ttttaaaat	aaaatgaagt	tttaaatcaa	tctaaagtat	atatgagtaa	acttggcttg
	10981	acagttacca	atgcttaata	agtggaggcc	ctatctcagc	gatctgtct	tttctgtcat
	11041	ccatagtgc	ctgactcccc	gtctgtttaga	taactacatg	acgggggggg	ttaccatctg
	11101	gcccccagtgc	tgcataatgata	ccgcgagacc	cacgcctacc	ggctcccgat	ttatcagcaa
	11161	taaaccagcc	agccggaaagg	gcccggcgca	gaagtggtcc	tgcacacttta	tccgcctcca
	11221	tccagtctat	taatttgtgc	cgggaaagcta	gagtaagttag	ttcgcggcatt	aatagttgc
	11281	gcaacgttgt	tgccattgtc	acaggcatcg	tggtgtcagc	ctcgtcgttt	ggtagggcgtt
10	11341	cattcagetc	cggttcccaa	cgatcaaggc	gaggtaatcg	atccccccatg	ttgtgcaaaaa
	11401	aagcgggttag	ctcccttcgg	cctccggatcg	ttgtcagaag	taagtggc	gcagtgttat
	11461	cactcatggt	tatggcagca	ctgcataatt	ctctactgt	catgcctatcc	gtaagatgt
	11521	tttctgtgac	tggtgagtag	tcaaccaagt	cattctgaga	atagtgtatg	cgccgaccga
	11581	gttgccttg	ccggcgctca	atacgggata	atacggcgcc	acatagcaga	actttaaaag
15	11641	tgctcatacat	tggaaaacgt	tcttcggggc	gaaaactctc	aaggatctta	ccgctgttg
	11701	gatccagttc	gatgttaaccc	actcgtgcac	ccaaactgatc	ttcagcatct	tttactttca
	11761	ccagcggttc	tgggtgagca	aaaacaggaa	ggccaaaatgc	cgccaaaaaag	ggaataagg
	11821	cgacacggaa	atgttgaata	ctcataactt	tcttttttca	atattattgt	agcattttatc
	11881	agggttatttgc	tctatgatgc	ggatataat	ttgtatgtat	tttagaaaaat	aaacaatata
20	11941	gggttcccgat	cacattttcc	ccaaaaaqtgc	cac		

2761 atgcttcaac taacatgtt gcttgcgggc gttcatgctc agaaaacaagg ttgggacaag
 2821 cacttccagg ctaacacagt cagaaaatcga aacgtactct caacagttcg cttaggcatg
 2881 gaagtttgc ggcattctgg ctacacaata acaaggaaag acttactcgt ggctgcaacc
 2941 ctactagtc aaaaattttt cacacatgtt tacgtcttgg gaaaattatg aggggatcgc
 5 3001 tctagagcga tccgggatct cggggaaagc gttggtgacc aaagggtgcct ttatcatca
 3061 cttaaaaat aaaaaacaat tactcagtc ctgttataag cagaattaa ttatgattga
 3121 tgctcacatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag aagttatatt
 3181 tgaacattat cttgattata ttattgataa taataaaaaac ttatcccta tccaagaagt
 3241 gatgcctatc attgggttgg aatgaacttga aaaaattag cttgaatac attactggta
 10 3301 aggttaacgc cattgtcago aaattgtatcc aagagaacca acttaaagct ttctgtacgg
 3361 aatgttaatt ctgcgttgc ctgagcactg atgaatcccc taatgattttt gttaaaaatc
 3421 attaagtttaa ggtggataca catcttgcata tgcatacccg gtaatgtgag ttagctcact
 3481 cattaggcac cccaggctt acactttat cttccggctc gtatgttgc tggatttgc
 3541 agccgataac aatttcacac agggaaacgc tatgaccatg attacgcacca gggcgcaatt
 15 3601 aaccctcaact aaagggaaaca aaagctggag ctccaccgcg gtggcggccg ctctagaact
 3661 agtgatccc cccggctgc ggaattcgat atcaagctt cgcataccgc tgacctcgag
 3721 ctgcagaaaa atgcagggtg gactatgaac tcacatccaa aggagcttga cctgataccct
 3781 gatttcttc aaacagggga aacaaccaa tcccaaaaaa cagctcagag agaaaccatc
 3841 actgatggct acagcacca ggtatgcata ggcacatcatc tctgtgaccc
 3901 gagcaaaatg atttctctc ccatgaatgg ttgccttccctcatgaa aaggcaattt
 3961 ccacactcac aatatgcac aaagacaatg agagaacaaat taatgtgctc ctccctaatt
 4021 tcaaattgt agtggcaag aggagaacaa aatctcaatg tctgagtagg ttttagtgc
 4081 tggataagag gctttgcact gtgagctcactc ctggacttca tattttttt gataaaaatg
 4141 gctttataa ctttcaggc tccgagttt tattcatgag actgttgggt tagggacaga
 4201 cccacaatga aatgcctggc atggaaagg gcagcagacg cttagctgac cttttcttgg
 4261 gacaaggcattc gtcacaaatc gtgtgacaaa acttattgtt ctgccttgc cagctgtgc
 4321 gggcaggccg atccattgc acctatccca ggtacatccctc caactgcac agattttgc
 4381 cttaactctt ctagaaatgt tctgcagatc gacatgcattt ctagatgttgc agataacatt
 4441 tactggaaag cacatctatc atcacaaaaaa gcagggcaaga ttttcagact ttctttgttgg
 4501 ctgaaaataga agcaaaaagac gtaattaaaaa acaaaatgaa aaaaaaaaaa tcagttgata
 4561 cctgtgtgt agacatccag caaaaaaaaaa ttatttgcac taccatcttgc tcttaagtcc
 4621 tcaagacttag caaggagaat gtagatttttcc acagtataat ttttttcaca aagggaaagga
 4681 gagaacaaaaa agaaaaatggc actgactaaa cttcagatgt tttttatggg aagtaatttgc
 4741 gcttaacaga gattgcgtt atctctatgt atgtctgtt gaaattatgtt gtactttttt
 4801 ccccccatttt taaatcaac agtgcatttcc agaggtcaga atggttttctt tactgtttgt
 4861 caattctattt atttcaatac agaacaatag cttctataac tggaaatataat ttgttattgt
 4921 atattatgt tgcctcctgc accatgaaaca ctcctccagc tgaatttccac aattccctctg
 4981 tcacatgcac ggcattaaatg ttattcatgg aagatcttgc aggaacactg caagttcata
 5041 tcataaaacat atttggaaattt gatgttttgc ttgcattgtt tggagctatg ttttgcgtt
 5101 tcctcagaaaa aaaaatgtt gataatggcatttgc tccacccat aaaaagatag attttaaat
 5161 tccagctata gggaaaggat gtcgtctgtt ctctacttca gtctcagtttgc tcccttcac
 5221 atqcatgtttt ctttatttttgc ctttatttttgc caagaaaaata ataggtcagc ttttttttgc
 5281 acttatgtcc tgcctcgttgc ggcctcagatg cacgttgcgtt atacaagaag gatcaaattt
 5341 aacagacttc tggctctgttgc cctacaacca tagtaataag cacaactaact aataatttgc
 5401 aattatgtttt tccatctcttca aggtttccat atttttctgt ttttttttttgc atcccattat
 5461 ctgggtttaa ctgaagcttca atggaaatgc agcaatatttgc cccagtttgc ttttttttgc
 5521 aacagtcttgc atggatttgc agaaggccgca gaaaacatc ttttttttttgc aaaaaaaaaa
 5581 taatattttgc tctccatttca atccaaaatgc gacatgttgc aactaaaatc taacccaaatc
 5641 ccattaaaatg atttctatgg tgcatttttttgc ctttttttttgc aagggaaatc ttttttttgc
 5701 cacaatttgc gctatatttgc ccccaagggtt cagccatgtt atcaatcatttgc ttttttttgc
 5761 ctttttttttgc ttttttttttgc acctatatttgc ctttttttttgc ttttttttttgc
 5821 aatataatttgc gacatgttgc ttttttttttgc ttttttttttgc ttttttttttgc
 5881 ctttttttttgc agtttttttttgc ggcctcagatg cgggggggggggggggggggggggggggggg
 5941 cgg
 55 6001 tcacttttgc aacgccttgc atggatgttgc tccgcaggc tccaggaaag gggctggat
 6061 gggctggccg tatttttttttgc aaaaatttttttgc gtttttttttgc ttttttttttgc
 6121 aaggccatgc caccatcttca agagatgttgc ttttttttttgc ttttttttttgc
 6181 gcctgaaatc cgaggacaca ggcgtatatttgc ttttttttttgc ttttttttttgc
 6241 ggggaggatatttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6301 ccaccaaggccatgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6361 cagccggccatgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6421 actcaggccatgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6481 tctacttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6541 tctgttgcacatgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6601 ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6661 ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6721 tcacatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc
 6781 tggacggccatgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc

5841 cgtacccgtgt ggtcagcgtc ctcacccgtcc tgcaccaggaa ctggctgaat ggcaaggagt
 6901 acaagtgc当地 aaagccctcc cagccccat cgagaaaaacc atctccaaag
 6961 ccaaaggc当地 gccccgagaa ccacagggtgt acacccgtcc cccatcccg gatgagctga
 7021 ccaagaacca ggtcagc当地 acctgc当地 tcaaaggctt ctatccc当地 gacatcgcc
 5 7081 tggagtggaa gagcaatggg cagccggagaa acaactacaa gaccacgc当地 cccgtgctgg
 7141 actcccgacgg ctccctt当地 ctctacagca agctcaccgt ggacaagagc aggtggcagc
 7201 aggggaacgt ct当地 catgc tccgtatgc atgaggctct gcaaccac tacacgc当地
 7261 agagccctc当地 cctgtctccg ggtaaaggc当地 cagaggccgaa gcttccat gagctgac
 7321 agccaccctc当地 ggttc当地 tccccc当地 caggaa aacccggc当地 gatcacctgc tetggagat
 10 7381 cattgccc当地 aaaaatggg tatttgatcc agcagaatgc aggccaggcc cctgtggg
 7441 tc当地 atatgc ggacagc当地 cgaccctccg ggatccctga gagattctct ggctccagct
 7501 cagggacaat ggc当地 accctt当地 actatc当地 gggccc当地 ggaagatgaa ggtgactact
 7561 actgttactc aactgacagc agtggat当地 atagggaggt gttcagc当地 gggacc
 7621 tgaccgtcc local agtgc当地 aaggctgccc cctcggtc当地 tctgttccca cccctct
 15 7681 aggagctca acccaacaag gccacactgg tggatctcat aagtgactcc taccgggg
 7741 cc当地 gacact ggc当地 ggaa gca gccc当地 gggggaggt gagacc
 7801 caccctccaa aacaaggc当地 aacaatgc当地 cggccaggc当地 ctacctgac
 7861 agcagtgaa gttcccaaaa agctacagc gccaggc当地 gcatgaaaggg agcacc
 7921 agaagacagc ggccctgca gaatgttccac egcggaggaa gggaaaggcc
 20 7981 ggggaggaaa cttc当地 gcca tgactccctc cgtgcccccc
 8041 gagggccctc tggcattgtc gcttccctc当地 cc当地 tccat
 8101 ttgc当地 tactgc cacagc当地 aataaaatcc
 8161 caagactgtc taagc当地 cttccccc当地 caccctt当地
 8221 caacttggc当地 actgctgccc atgagaa
 25 8281 attacatatg cccatagatcc tgattaacag gtgtt当地
 8341 cccacattat cccattgc当地 cccctcgagg
 8401 gagtc当地 tatttgc当地 ac
 8461 ggc当地 ttaccc aacttaatgc
 8521 gaaggaggccc
 30 8581 ttgc当地 taatgttgc
 8641 ttaaccata gggc当地 gaaatcc
 8701 gggtt当地 gagtgc tggc当地
 8761 tcaaaaggc当地
 8821 gatgtt当地 tacccttaact
 8881 tc当地 aggtc当地
 8941 acttactca
 9001 cgataaaaaa
 9061 ataaaataga
 9121 gggtt当地 atcaag
 40 9181 gc当地 acttgc当地
 9241 aggataataa
 9301 agtgc当地 atataataa
 9361 aggctc当地
 9421 actccctgta
 9481 actaaccac
 9541 cc当地 gagc当地
 9601 ggaactact
 9661 catc
 50 9721 ataatcaaca
 9781 agtggagg
 9841 gttatcc
 9901 gtgc当地
 9961 cgg
 10021 tgc当地
 10081 tgc当地
 10141 ata
 10201 cc当地
 10261 gtc当地
 10321 gaag
 10381 ttcc
 10441 tgc当地
 10501 ggc当地
 10561 tgg
 10621 tctt
 10681 tgc当地
 10741 cc当地
 10801 ctc当地
 10861 gtt
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65

5	10921	aaaaatgaag	ttttaatca	atctaaagta	tatatgagta	aacctgggtct	gacagttacc
	10981	aatgcttaat	cagtggagca	cctatctcag	cgatctgtot	atttcgttca	tccatagttg
	11041	cctgactccc	cgtcgtag	ataactacga	tacggggaggc	cttaccatct	ggccccagtg
	11101	ctgcaatgt	accgcggagac	ccacgctcac	cggtccca	tttatcagca	ataaacccagc
	11161	cagccggaa	ggccggagcg	agaagttggc	ctgcaactt	atccgccttc	atccgactta
	11221	ttatattgtt	ccggaaagct	agagaatgt	gttcggcagt	taatagttt	cgcaacgttg
	11281	ttgccccattgc	tacggccatc	gtgggtgtac	gctcgtcgtt	ttgtatggct	tcattcagct
10	11341	ccgggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	aaagcgttta
	11401	gctccattcgg	tcctccgatc	gttgcagaaa	gtaagttggc	cgcaagtgtt	tcactcatgg
	11461	ttatggcagc	actgcataat	tctttaactg	tcatgccatc	cgtaagatgc	ttttctgtga
	11521	ctggtgagta	ctcaaccac	tcattctgag	aatagtgtat	gccccgcaccc	agttgtcttt
	11581	gccccgggtc	aatacgggat	aataccggc	cacatcgac	aactttaaa	gtgtcatca
	11641	ttggaaaacg	ttcttcgggg	cgaaaactct	caaggatctt	accgtgtt	agatccagtt
15	11701	cgatgtaa	cactcgatc	cccaactat	cttcagatc	ttttacttc	accaggctt
	11761	ctgggtgagc	aaaaacacgg	aggccaaatg	ccgcaaaaaaaa	gggaaaaaagg	gccccacccg
	11821	aatgttgaat	actcatactc	ttcccttttc	aatattattt	aagcatttt	cagggttatt
	11881	gtctcatgag	cggatacata	tttgaatgt	tttagaaaaaa	taaacaaata	ggggttccgc
	11941	gcacatttcc	ccgaaaatgt	ccac			
20	SEQ ID NO: 99 (pTnMCS (CHOvep-prepro-HCPro-Lys-CPA))						
	1	ctgacgcgc	ctgtacggc	gcattaagcg	cgccgggtgt	gttgggttacg	cgccagcgta
	61	ccgctacact	tgccagcgcc	ctagcgcccg	ctcccttcgc	tttcttccct	tccttttctcg
25	121	ccacgttgc	ccgcatacga	ttggcttattg	gccattgtatc	acgttgtatc	catactcataa
	181	tatgtacatt	tatattggct	catgttccaa	attaccgc	tgttgacatt	gattatttgac
	241	tagttattaa	tagtaatcaa	ttacgggggt	attagtgtatc	agccatata	tggagttccg
	301	cgttacataa	cttacggtaa	atggcccgcc	tggctgaccc	cccaacgacc	cccccccat
	361	gacgtcaata	atgacgtat	tttccatagt	aaagccaa	gggactttc	attgacgt
	421	atgggtggag	tatattacgt	aaactgc	cttggcgat	catcaagtgt	atcatatgc
30	481	aagtacgc	ccatttgc	tcaatgc	taatggccc	gcctggcatt	atgeccagta
	541	catgaccctt	tggacttcc	ctacttgc	gtacatctac	gttattgtca	tcgttattac
	601	catggtgat	cggttttggc	agtacatcaa	tggcggttgc	tagcggttt	actcacgggg
	661	atttccaa	ctccacccca	ttgacgttca	ttggagtttgc	tttggcacc	aaaatcaacg
35	721	ggactttca	aaatgtcgt	acaactccgc	cccaattgc	caaatatggcg	gtaggcggt
	781	acgggtggag	gtctatataa	gcagagctcg	tttagtgcac	cgtcagatcg	cctggagacg
	841	ccatccacgc	tgttttgc	tocatagaag	acaccggac	cgatccacgc	tccggcccg
	901	gaaacgggtc	attggaaacgc	ggattccccc	tgccaaaggt	gacgtaa	ccgcctata
	961	actctatagg	cacaccctt	ttggcttctt	gcatgtata	ctgtttttgg	tttggggct
40	1021	atacacccccc	gttcttctt	gtataggt	atgttatagc	ttageotata	ggtgtgggtt
	1081	attgaccatt	attgaccact	cccttattgg	tgacgatact	ttccattact	aatccatataac
	1141	atggcttctt	ggccacaacta	tcttcttattgg	ctatatgc	atactctgtc	tttccatagac
	1201	tgcacacggac	tctgttattt	ttacaggatgg	ggtccattt	atttatttaca	aatttacat
	1261	tacaacaacg	ccgtcccccc	tgcccg	ttttatttttt	catagcgttgc	gatctccacg
45	1321	cgaaatctcg	gtacgtgtc	eggacatggg	cttttctccg	gtagcggcg	agcttccac
	1381	tccgagccct	ggtcccatgc	ctccagccgc	tcatggcgc	tcggcagctc	tttgcctcta
	1441	acagtggagg	ccagactttag	gcacagcaca	atgcccacca	ccaccagtgt	gccgcacaag
	1501	gcccgtggcg	tagggatgt	gtctggaaat	gagcgtggag	atggggctcg	cacggctgac
	1561	gcagatggaa	gacttaaggc	agcggcagaa	gaagatgc	gcacgttgc	tgttgtattc
	1621	tgataaaggt	cagaggtaac	ttccgttgc	gtgtgttgc	ccgttgggggg	cagtgtatgc
	1681	tgacgtatc	tcgttgc	cgccgcgc	accagacata	atagtgaca	gactaaacaga
	1741	ctgttctt	ccatgggtc	tttctcgat	cacccgtcg	ccatgtcgca	actcgatatt
	1801	ttacacgact	ctcttacca	attctgc	cccc	taaaacgact	caacagctt
	1861	acgttggctt	ggcacgcatt	acttgc	actgt	taaaactctca	ctcttaccga
	1921	aacctgc	ccaaagcg	aacaaacat	aacatcaa	gaatcgacc	attgttaggt
50	1981	aatcgtcacc	ttccacaaaga	gcgactcg	gtataccgtt	ggcatgttgc	ctttatctgt
	2041	tcgggcata	cgatccccc	tgtacttgc	gactggc	atattcgta	gcaaaaaacga
	2101	ttatattgtat	tgcgagctt	agtcgacta	cacggcgt	ctgttactt	ttatgaaaa
	2161	gggttcccg	tttccatgc	atgttca	aaagctcat	accatatttct	agccgacatt
	2221	ggggatcc	taccggat	ccacaccc	cttacgttca	gtgtatgttgc	ctttaaatgt
	2281	ccatggtata	aatccgttgc	gaagctgg	ttgtacttgc	taatgtcgat	aaggagaaaa
	2341	gtacaatatg	cagacctagg	agcggaaaac	tggaaaccta	tcgcaactt	acatgatatg
	2401	tcatctatgc	actccaaagac	tttaggtat	aagaggctg	ctaaaagca	tccatctca
	2461	tgcacacgg	tattgtataa	atctcgctc	aaaggccgaa	aaaatcagcg	ctcgacacccg
	2521	actcattgtc	accacccgc	acttacaaat	tactcagcgt	cgccaaagga	gccccatgggtt
	2581	ctagcaacta	acttacatgt	tggaaatcg	acacccaa	aacttgcata	tatctatcg
	2641	aagcgaaatgc	agattgtt	aaaccccg	gacttgc	tttgccttgc	cgggacttgc
	2701	tcacgcccata	ggccgaaacgg	cagctc	tttgcata	ttatgtcgat	aatccatctcg
	2761	atgcttcaac	taacatgttgc	gttgcggc	tttgcata	aaagaaacaa	ttgggacaa

2821 cacttccagg ctaacacagt cagaatcgaa aacgtactct caacagttcg cttaggcatg
 2881 gaagtttgc ggcattctgg ctacacaata acaaggaaag acttaactcggt ggctgcAAC
 2941 ctactagtc aaaaatttatt cacacatgg tacgctttgg gggaaattatg aggggatcg
 3001 tctagagcgta tcggggatct cggggaaagc gttggtgacc aaagggtgcct ttatcatca
 5 3061 ctttaaaaat aaaaacaat tactcagtgc ctgttataag cagaattaa ttatgattga
 3121 tgcctacatc acaacaaaaa ctgatTTaa aatgggtgg tctgccttag aaagtatatt
 3181 tgaacattat ctgttattata ttattgataa taataaaaaac cttatcccta tccaagaagt
 3241 gatgcctatc attgggtgg aatgacttga aaaaaattag ccttgaatac attactggta
 3301 aggttaaacgc cattgtcagg aatTTatgc aagagaacca acttaaagct ttccgtacgg
 10 3361 aatgttaatt ctgttgcgg ctgagactg atgaatcccc taatgatTT ggtaaaaatc
 3421 attaagttaa ggtggataca catTTgtca tatgtcccg gtaatgtgag ttagctca
 3481 cattaggcac cccaggctt acactttatg cttccggctc gtatgtgtg tggaaatttg
 3541 agcggataac aatttcacac agggaaacagc tatgaccatg attacgcca ggcgcatt
 3601 aaccctcaact aaagggaaaca aaagctggag ctccaccggc gtggggccg ctctagaact
 15 3661 agtggatccc cccggctgc ggaattcgat atcaagctta tcgataccgc tgacctcgag
 3721 ctgcagaaaaa atgcccagggt gactatgaa tcacatccaa aggagttga cctgatacc
 3781 gattttcttcc aaacaggggaa aacaacacaaa tcccacaaaaa cagtcagag agaaaccatc
 3841 actgtatgc acagcacca ggtatggat ggcattccat tcgacattca tctgtgacct
 3901 gagaaaaatg atttcttctt ccatgtatgg ttgttctt ccctcatgaa aaggcaatt
 20 3961 ccacactcac aatatgcgac aaagacaaaac agagaacaat taatgtgc ttccttaatg
 4021 tcaaaaattgt agtggcaag aggagaacaa aatctcaatg tctgatggg ttttagtgc
 4081 tggataagag gcttgcacct gtgagctcac ctggacttca tatttttg gataaaaaatg
 4141 gttttatataa ctttcagggtc tccgagtttca tattcatgag actgttgggt tagggacaga
 4201 ccccaatgaa atggctggc atggaaagg gcgcggcgtc cttatgtgac cttttttgg
 25 4261 gacaaggattt gtcaacaaat gtgtggataaa actatTTgtc tgccttgc cagctgtgct
 4321 gggcaggccg atccatttgc accatccca ggttacccca caactgcgaa aagattttg
 4381 cttactctct ctagaaagct tctgcagact gacatgcatt tcataggtag agataacatt
 4441 tactggaaag cacatctatc atcacaaaaa gcaggcaaga tttttagact ttcttagtg
 4501 ctgaaataga agcaaaagac gtaattaaaa acaaaatgaa aaaaaaaaaa tcagttgata
 4561 cctgtgggtg agacatcccg caaaaaaaat ttatTTgcac taccatcttgc tcttaagtcc
 4621 tcagacttgc caaggaaat gttagatttcc acatgtatataa tggggatcaca aaagggagg
 4681 gagaaaaaaa agaaaaatggc actgactaaat cttcagctgt tggatggaa aagttatct
 4741 gcttaacaga gattgcaggta atcttctatgt atgttgcgtt gattttatgtt gtaactttt
 4801 cccccatttt taaatcaaaat agtgcTTTAC agaggatcaga atgggttctt tactgtttgt
 4861 caatttttattt atttcaatac agaacaatac cttctataac tgaatataat ttgttattgt
 4921 atattatgtat tgccttcgc accatgtatc cttccctccgc tgaatttcac aattccctcg
 4981 tcatatgcgc ggcattaaat ttattatgtt gatgttttgc aggaacactg caagttcata
 5041 tcataaaacat atttggaaat gatgttttgc ttgcatttgc tggagctatg ttttgcgt
 5101 tcctcggaaaaaaaatggat taaaatggat tcaacccatc aaaaagatag attaaatat
 5161 tccagctata gggaaaaggg tgcgtctgc cttcacttca gtcctagtt gtccttcac
 5221 atgcattgtt ctttatttttgc cttatTTTGC caagaaaaata ataggcactg tcttttttgc
 5281 acttattgtcc tgccttagcat ggctcagatg cacgttgcgtt atacaagaag gatcaatga
 5341 aacagacttc tggctgtta cttacaacca tagtataag cacactaact aataattgt
 5401 aattatgtttt tccatcttca aggttccatg atttttctgt ttcttttttttgc atcccttatt
 5461 ctgggtgttgc ctgaaatgcg atggaaatcg agcaatattt cccagtttgc tctccatcc
 5521 aacatgttgc atggatgttgc agaaggccg gaaaacacat tggggatccg aattaaaaac
 5581 taatatttgc tttccatttca atccaaaaatg gacctatgt aactaaaaatc taacccatc
 5641 ccattaaatg atttctatgg tgcattttttgc caacccatc tttttttttttttttttttttt
 5701 cacaatttcg gctatatttttgc cttccatccgc tggggatccg aattaaaaatc
 5761 cttcttcgtg tgggtgtttt accttatatttgc tttttttttttttttttttttttttttttt
 5821 aatatttttc gacatgtat tt
 5881 ctttgcac agtttgcgtt gggccaggatc cggggatccg cttccatccgc tggggatccg
 5941 cgg
 6001 tcaacttcg aaacccctgg atggatgggg tccggccggc tccgggggggggggggggggggg
 6061 ggggtggccgg ttt
 6121 aaggcagattt cccatcttca agatgttgc ttttttttttttttttttttttttttttttttt
 6181 gcctgaaagc cggggatccgc gggggatccgc ttttttttttttttttttttttttttttttt
 6241 gggggatccgc ttt
 6301 cccatccatccgc ttt
 6361 cggggccctt gggccatccgc ttt
 6421 actcggccgc cttccatccgc gggccatccgc ttttttttttttttttttttttttttttttt
 6481 ttt
 6541 ttt
 6601 ttt
 6661 ctt
 6721 ttt
 6781 ttt
 6841 ctt

6901 acaagtgcac ggtctccaaac aaagccctcc cagccccat cgagaaaacc atctccaaag
 6961 ccaaaggcga gccccgagaa ccacagggtg acaccctgc cccatcccg gatgagctga
 7021 ccaagaacca ggtcagcctg acctgectgg tcaaaggctt ctatccage gacategcgg
 7081 tggagttggga gagaatggg cagccggaga acaactaca gaccacgc cccgtctgg
 5 7141 actccgcacgg ctcccttc ctctacagca agctcaccgt ggacaagagc aggtggcagc
 7201 agggaaacgt ctctcatgc tccgtatgc atgaggctc gcacaaccac tacacgcaga
 7261 agagctctc cctgtctcg ggtaaagcgc cagagccaa aaagcttcc tatgagctga
 7321 cacaggccacc ctccgtgtca tggtccccag gacaaacgc caggatcacc tgctctggag
 7381 atgcattgtcc agaaaaatat ttgtattggt accagcaga gtcagggccag gcccctgtgg
 10 7441 tggtcatcta tgaggacagc aaaaacccctt cgggatccc tgagagattc ttcgttcca
 7501 gctcaggccat aatggccacc ttgactatca gtggggccca ggtggaaagat gaaggtgact
 7561 actactgtta ctcaactgtac agcagtgggt atcataggga ggtgttcagc ggaggggacca
 7621 agctgaccgt cctaggcgtc cccaaaggctg cccctcggt cactctgttc ccacccctct
 7681 ctgaggagct tcaagccaaac aaggccacac ttgtgtgtct cataagtgcac tccctcccg
 15 7741 gagccgtgac agtggcctgg aaggcagata gcagcccgtaaaggccgtt caaggccggg gtggagacca
 7801 ccacaccctc caaacaacagc aacaacaatg acggcccgac cagctacccctc agccgtacgc
 7861 ttgagcagtg gaagttcccaaaaatcata gtcgtccagg cagcatgaa gggagcaccg
 7921 ttggagaagac agtggcccccgtc gcaaatgtt caccggcggag ggagggaagg gccccttttg
 7981 aagggggagg aaacttcgcttccatgactcc ttcgtgtcccccgcacgggaaactgtatgt
 20 8041 gcagagggcc ctctgcccattt gtcgttccctt ctggcccttcc ttgtcaactctt gaatgtggct
 8101 tctttgtctac tgccacagca agaaataaaaaa ttcacacatc taaatgggtt tcttgagatt
 8161 tttcaagatgt cgttaagcac attcccttccctt cagcaccctt tgctgcaggc cagtgcagg
 8221 caccacacttgc gtcactgtcc cccatgagaaatccatgtt caatattttt caaagcaaaa
 8281 ttggatcatat atggccatgttccatgtt ctcgttattttt ctcgttgcgtt gtttttcgtt
 25 8341 tcaccacat ttccttcattt ctcgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8401 agtgcgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8461 cttggcgatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8521 agcgaagagg cccgcaccgc ttcgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8581 aaattgttaag cgttaatatttttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8641 ttttttaacca ataggccaaatccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8701 taggttggatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8761 acgtccaaagg gggaaaaacc gtcgtatcagg gtcgtatcagg gtcgtatcagg gtcgtatcagg
 8821 caagatgtgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8881 tgctcagggttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 8941 aaaacttacttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9001 tgccgataatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9061 taaaatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9121 ttgggttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9181 taagcacttgcgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 40 9241 agcaggatataatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9301 ggtgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9361 gtaagcttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9421 atcactccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9481 aaaactaacc aaccccttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 45 9541 aatcccgatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9601 cttggatataatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9661 catcatcaccgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9721 ataataatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9781 ttttagtggatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9841 attgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9901 ggggttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 9961 agtccggaaaatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10021 gtttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 55 10081 ggctcgccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10141 gggataacgcgatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10201 aggcgcgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10261 gacgcgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10321 ctggaaatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10381 ctttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 60 10441 cgggttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10501 gtcgtccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10561 cactggcaccgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10621 agtccgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10681 ctctgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10741 ccaccgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10801 gatccgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10861 cacgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt
 10921 attaaaaatccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgttccatgtt

5	10981	accaatgctt	aatcagtgag	gcacccatct	cagcgatctg	tcttatttcgt	tcatccatag
	11041	ttgcctgact	ccccgtcgtg	tagataacta	cgatacggga	gggcttacca	tctggcccca
	11101	gtgctgcaat	gatacccgca	gaccacgct	caccggctcc	agattttatca	gcaataaaacc
	11161	agccagccgg	aaggcccgag	cgcagaatgt	gtcctgcacac	tttatccggc	tccatccagt
	11221	ctattaattt	ttgcgggaa	gctagagatg	gtatgttgc	agttataagt	ttgcgcacacg
	11281	ttgttgcatt	tgttacaggc	atcggtgtt	cacgctgc	gtttgtatg	gttccatcca
	11341	gtccgggtc	ccaacgatca	agggatgtt	catgtatccc	catgttgc	aaaaaaacgg
	11401	ttagctccct	cggttcccg	atcggtgtt	gaagtaatgtt	ggccgcagt	tttactactca
	11461	tggttatggc	agcaactgcat	aattcttta	ctgtcatgcc	atccgtaa	tgctttctg
	11521	tgactgggtg	gtactcaacc	aagtcttct	gagaatagtg	tatgcggcga	ccgagttgtt
	11581	tttgcggcgc	gtcaatacgg	gataataccg	cgccacatag	cagaacttta	aaagtgtc
	11641	tcattggaa	acgttcttc	ggccggaaaac	tctcaaggat	cttaccgc	ttgagatcca
	11701	gttcgatgt	accactgt	gcacccaaact	gatcttc	atctttact	ttcaccagcg
	11761	tttctgggt	agcaaaaac	ggaaggcaaa	atggccaaa	aaaggaaata	agggcacac
	11821	ggaaatgtt	aataactata	ctcttc	ttcaatatta	ttgaagcatt	tatcagggtt
	11881	attgtctcat	gagcggatac	atatttgaat	gtatgtt	aaataaaacaa	atagggtt
	11941	cgcgcacatt	ccccggaaa	gtgcac			
	SEQ ID NO:100	(pTnMCS (CMV-prepro-HCPro-CPA))					
20	1	ctgacgcgc	ctgtacgcgc	gcataacgc	cggggggtgt	ggtggttacg	cgacgcgtga
	61	ccgttacact	tgccagcgc	ctagcgc	ctcccttgc	tttcttc	tcctttctcg
	121	ccacgttcgc	cgccatcaga	ttggatttgc	gcattgc	acgttgc	tatcatataa
	181	tatgtacatt	tatattggct	catgtccaa	attacggca	tggttgc	attttatgg
	241	tagttattaa	tagaatcaa	ttacggggtc	attagttcat	agccatata	tggagttccg
	301	cgtttacataa	cttacggtaa	atggccgc	tggtgc	cccaacgc	ccgcgcatt
	361	gacgtcaata	atgacgtat	ttccatagt	aacgcaata	gggactttt	attgacgtca
	421	atgggtggag	tatattacgt	aaactgc	cttggc	catcaagt	atcatatgc
	481	aagtacgc	cctatgc	taatgcac	taatggcc	gcctgt	atgcccagta
	541	catgaccta	tgggactt	ctacttgc	gtatcatc	gtattagtc	tcgttatt
	601	catggtgat	cggttttgc	agtacatca	tggcgttga	tagcggtt	actcacggg
	661	atttccaat	ctccacccca	ttgacgtca	tgggagttt	ttttgc	aaaatcaacg
	721	ggactttca	aaatgtcgta	acaactccgc	ccattgc	caaatggc	gtaggcgt
	781	acgggtggag	gtctatataa	gcagagct	tttagtga	cgtagat	cctggagacg
	841	ccatccacgc	tggttttgc	tccatagaag	acacccggac	cgatccagg	ccgcgcgc
	901	ggaaacgggt	atggacatgc	ggatccccc	tccaaagat	gacgttga	ccgcctatag
	961	actctatagg	cacacccctt	ttggctt	gtatgtata	ctgttttgc	tttgggc
	1021	atacacccccc	gttcc	gtatgttgc	atgttata	ttagctata	ggtgtgtt
	1081	attgaccatt	attgaccact	ccccttattt	tgcgtact	ttccattact	aatccataac
	1141	atggctctt	gccacaacta	tcttattt	ctatgtca	atactctgt	tttgcagagac
	1201	tgacacggac	tctgtt	tacaggatgg	ggtcc	attatttaca	aattcacata
	1261	tacaacaacg	ccgtccccc	tgccgc	agt	catagcgt	gatctccacg
	1321	cgatctcg	gtacgttgc	ggacatgg	cttcttcc	tgacgggg	agttccaca
	1381	tccgagccct	ggtcc	ctccacgc	tcatggc	tcggcagtc	cttgcctata
	1441	acagtgagg	ccagactt	gcacagcaca	atggcc	ccaccatgt	gcgcacaa
	1501	gccgtggcg	taggtatgt	gtct	gaaat	attgggc	tacggcgtac
	1561	gcagatggaa	gacttaaggc	agcggc	gaaat	gcagct	tggttatt
	1621	tgataagat	cgaggtaac	tccgttgc	gtgt	cggtggagg	cagtgtat
	1681	tgagcgtac	tctgttgc	cgcg	cgcc	atagctg	gactaaacaga
	1741	ctgttcc	ccatgggt	tttctgc	atccgt	ccatgt	actcgatatt
	1801	ttcacacgt	cttta	atcttgc	ccc	taaaacgact	caacacgtt
	1861	acgttggctt	gccc	acttgc	at	cttcc	acttggcg
	1921	aacctccaa	ccaaacgc	aaacaa	actt	gaatc	acttggatgt
	1981	aatcgtcacc	tccacaa	agcgtc	gtat	gcac	tttgc
	2041	tcggc	cgat	tgtactt	gact	ggcat	tttgc
	2101	tttgc	gat	gcact	ggt	atatt	ttatgagaaa
	2161	gggttcc	tttca	atgttca	actt	ctgttact	ttatgagaaa
	2221	gcg	tc	ccacaccc	gc	accat	tttgc
	2281	ccatgtata	aatccgtt	gaag	ttgt	taat	tttgc
	2341	gtacaatat	cgac	ccgtt	actt	tcatgt	atatgtat
	2401	tcatctgt	act	ccgtt	actt	ctaaa	acaaatccaa
	2461	tgccaaat	tatt	tttgc	at	aaaat	ccatctt
	2521	actcatttgc	acc	ccgtt	cc	ccgc	ccgcacacg
	2581	ctagcacta	actt	ccgtt	at	ccg	ccgcgg
	2641	aagcgtatgc	agat	ccgtt	actt	tcatgt	atatgtat
	2701	ctacgcata	ggc	ccgtt	actt	ctaaa	acaaatccaa
	2761	atgcttca	taac	ccgtt	at	aaaat	ccatctt
	2821	cacttccagg	ctaa	ccgtt	cc	ccgc	ccgcacacg
	2881	gaagtttgc	ggc	ccgtt	at	ccg	ccgcgg

2941 ctactagctc aaaattttt cacacatgg t acgctttgg g gaaattatg agggatcgc
 3001 tctagagcga tcgggatct cgggaaaagc gttggtgc a aaggtgcct tttatcatca
 3061 ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa ttatgattga
 3121 tgcctacatc acaacaaaaa ctgatttaac aaatggtgg tctgccttag aaagtatatt
 5 3181 tgaacattat ctgattata ttattgataa taataaaaac cttatcccta tccaagaagt
 3241 gatgcctatc atgggttgg a tgaacttga aaaaattag cttgaatac attactggta
 3301 aggtaaacgc cattgtcage aaattgtatcc a aagagaacca acttaaagct ttctgacgg
 3361 aatgttaatt ctgttgc a ctgacactg atgaatcccc taatgattttt ggtaaaaatc
 3421 attaagttaa ggttgcatac catctgtca t atgatccg gtaatgtgag ttatgtcact
 10 3481 cattaggcac cccaggctt acactttatg cttccggc t gatgttgg t gggattgtg
 3541 agcggataac a a t t c a c a c a g g a a a c a c a g t a t g a c c a t g a t t a c c a a g c g c g c a a t t
 3601 aaccctca c a a g g g a a c a a a a g c t g g a g c t c a c c a c c g e g g t g c g g c c g c t c t a g a a c t
 3661 agtggatccc c c c g g g c t g c a g g a a t t c g a t a t c a g t t a c g a t a c c g c t g a c c t c g a g
 3721 catcagattt gctattggcc attgcatacg ttgtatccat atcataatat gtacatttat
 15 3781 attggctcat g t c c a a c a t t a c c g c t a t g t a t g t a t t a c t a g t a t t a a t a g
 3841 tatacaattt c g g g g t c a t t a c t a g t a c t a t a t g g a g t t c c g c t a t a c t a a c t t
 3901 acggtaatg g c c g c c t g g c t a g c a t t a c t a g t a a c g a c c c c c g c a t t g a c g t a a t a g t
 3961 acgtatgttc ccatagtaac g c c a a t a g g g a c t t c c a t t a c g t c a a t g g t g g a g t a t
 20 4021 ttacggtaaa ctgcccactt g g c a g t a c a t a c a g t g t a t c a t g t c a a g t g c a a g t a c g c c c c t
 4081 attgacgtca atgacggtaa atggccgc t g g c a t t a t g a t t a c a t g a t t a t g g c a c c t t a t g g
 4141 gacttccca ctggcagta catctacgtt ttagtcatcg ctattaccat ggtgatgcgg
 4201 ttttgcgt acatcaatgg g c g t g g a t g c g t a c g t t t g a c a c g g g a t t t c c a g t c t t
 4261 cacccttgg agtcaatgg g a g t t t g t t t g g c a c c a a a a t c a a c g g g a c t t c c a a a a a
 4321 tgcgttaaca actccgc cccattt a t t g a c g c a a t t g g c g g t a a g c g c t g a c g g g a g g t c
 25 4381 tatataagca gagctcg tttt a g t g a a c c g t c a g a t e c c t g a g a c g c c a t c a c g c t g t
 4441 tttgacccctt atagaagaca c c c g g a c c g a c c g c t c c a g g t g a c t a a c t a a c a c t a a c t
 4501 ggaacgcgga ttcccccgtgc a a g a g t g a c g t a a c t a a c a t g a c t a a c t a a c a c g c a c
 4561 acccccttgg ctcttatgca t g c t a t a c t g t t t t g g c t a t a c t a a c c c c c g c t
 4621 tcctttagtct atagggtatg g t a t g a c g t t a c t a a c a t g t a t g t a t g g g t t a t t a c t a
 30 4681 gaccactccc ctatggtga c g a t a c t t c c a t t a c t a a t t a c t a a c a t g t a t t t g c c
 4741 acaactatct ctatggtca t a t g c t a a t a c t a a t t a c t a a c a t g t a a c t a a c t a a c t
 4801 gtatttttac agatgggg t c c c a t t t t a t t a c a a a t t a c a t a t a c a a c a c g c c g
 4861 tccccctgtc cccgagttt t a t t a a a c a t a c a g t g g g a t c c a c a c g c g a c t t c c a c
 4921 cgtttccgg acatgggctc t t c t c g g t a g c g g g a g t t c c a c a t t c c a g c c t g g t
 4981 cccatgcctc c a g c g g c t c a t g g c t c t t c t c t a a c a t g g a g g g c c a
 5041 gacttaggca c a g c a c a t g g c a c c a c c a c g t g t c c t c a a g g g c c g t g g c g g t a g
 5101 ggtatgtgtc t g g a a t t a a g g a g a g a g a g a g a g a g a g a g a g a g a g a g a g a g a g
 5161 ttaaggcagc g g c a g
 5221 aggttaactcc c g t t g c g g t g t g t a a c c g g a g a g a g a g a g a g a g a g a g a g a g a g
 5281 ttgctgcgc g c g c g c c a c c a g a c a t a a t a a t a a t a a t a a t a a t a a t a a t a a t a
 5341 tgggtctttt ctgcagtca c g t c g g a t c a t c a t c a t c a t c a t c a t c a t c a t c a t c a t
 5401 gtgttacact atatatctaa t a t t a a t a t t a a t t a a t a a t a a t a a t a a t a a t a a t a a t a
 5461 atgaattttt c a c a g g a t
 45 5521 tcggctgcgc c a g a g c g a a g g a t
 5581 gtaaagccgg g g g g g t c c t t c c t t c c t t c c t t c c t t c c t t c c t t c c t t c c t t c c
 5641 gcctggatga g c t g g g t c c c a g g g t c c a g g g g g g g a g g g g g g g g g g g g g g g g g g
 5701 aaaagcaaaa ttgatggtgg g a c a a c a g a c a t g t c a c c c g g t a a a g g a g g a g g a g g a g g
 5761 atctcaagag atgattcaaa a a a c a c g t t a a t c t g a a a t g a a t a g c c t g a a a g a g c c g a g
 5821 gacacagccg t a t a t a c t g t a a a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c
 5881 ccccccaattt g g g g c c a g g g a a c c c t t g t c c a c c a c c a c c a c c a c c a c c a c c
 5941 tcggctttcc c c t t g g c a c c a g g a c c a c c a c c a c c a c c a c c a c c a c c a c c a c c
 6001 tgcctggatca aggactactt t c c c c c a a c c a a c c a a c c a a c c a a c c a a c c a a c c a a c
 6061 accagggcg t g c a c a c c t t c a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c
 6121 aacgtggatca c c t g c c c t t c a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c
 6181 cacaagccca g c a a c a c c a a a g g t g a c a a g a a a g g t g a c a g a c c a a c a c a c a c
 6241 cacacatgcc c a c c g t g c c c a g c a c t g a a a t
 6301 cccccaaaaac c a c a g g a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c a c
 6361 gtggacgtca g c a c a g a a g c t g a a g g t t c a a c t g t a c g t g g a g g g g g g g g g g g g
 6421 gtgcataatg c c a a g a a a a g g c e g g g a g g a c g t a c a c a c a c a c a c a c a c a c a c
 6481 a g c t c t c a c c t t g t c a c c a c c a c a c a c a c a c a c a c a c a c a c a c a c a c a c
 6541 tccacaacaag c c c t c c a c a g a
 6601 cgagaaccac aggtgtacac c c t g c c c c c a
 6661 agcctgaccc g c c t g g t c a a a g g t t c t a a a c c a c g c a c a c a c a c a c a c a c a c
 6721 aatggccagc c g g a g a a c a
 6781 ttcttccttct acagcaagct c a c c g t g g a c a a a g g a c g g a g g a g g a g g a g g a g g a g g
 6841 tcatgtcgc t g a t g c a t g a g g a g g a g g a g g a g g a g g a g g a g g a g g a g g a g g a g g
 6901 tctccggta a a g c c c a g a g c a a g c t t c c t a t g a g c a c a c a g c c a c c t c c o t g
 6961 tcagtgtccc c a g g a c a a a c a

5	11101	cataatttctc	ttactgtcat	gccatccgta	agatgtttt	ctgtgactgg	tgagtactca
	11161	accaagtcat	tctgagaata	gtgtatcgccg	cgaccgagg	gctcttgc	ggcgctaaata
	11221	cgggataata	ccgcgcacaca	tagcagaact	ttaaaagtgc	tcatcattgg	aaaacgttct
	11281	tcggggcgaa	aactctcaag	gatcttacccg	ctgttgagat	ccagttcgat	gtaaaccact
	11341	cgtgcacccaa	actgatcttc	agcattcttt	actttccaca	gctttctgg	gtgagaaaaaa
	11401	acagggaaaggc	aaaatgcgc	aaaaaaggga	ataaggggca	cacggaaatg	ttgaataactc
10	11461	atacttcc	ttttcaata	ttatgttgc	atttacccg	gttattgtct	catgagcgga
	11521	catatatttgc	aatgtat	aaaaaaataaa	caaataagg	ttccgcgcac	atttccccga
	11581	aaagtgcac					
15	SEQ ID NO:101 (pTnMCS (CMV-prepro-HCPro-Lys-CPA))						
	1	ctgacgcgc	ctgtacggc	gcattaaagcg	cgccgggtgt	ggtggttacg	cgcaagcggtga
	61	ccgcctact	tgccagcgcc	ctagecccg	ctcccttcgc	ttttttccct	tcctttctcg
20	121	ccacgttgc	cggcatacga	ttggatgttgc	gcattgtat	acgttgtatc	cataataaa
	181	ttatgttgc	ttatgttgc	catgttcaac	attaccgc	ttgttgacatt	gattattgac
	241	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
25	301	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	361	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	421	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	481	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
30	541	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	601	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	661	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	721	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
35	781	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	841	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	901	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	961	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
40	1021	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1081	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1141	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1201	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
45	1261	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1321	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1381	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1441	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
50	1501	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1561	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1621	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1681	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1741	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1801	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
55	1861	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1921	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	1981	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2041	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
60	2101	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2161	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2221	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2281	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2341	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2401	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2461	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2521	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2581	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
65	2641	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2701	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2761	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2821	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2881	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	2941	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	3001	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	3061	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	3121	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	3181	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	3241	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	3301	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc
	3361	ttatgttgc	ttatgttgc	ttatgttgc	ttatgttgc	ttgttgacatt	ttggatgttgc

3421 attaagttaa ggtggataca catottgtca tatgatccc gtaatgttag ttagctact
 3481 cattaggcac cccaggctt acactttatg cttccggctc gtagttgtg tgaattgtg
 3541 agcgataac aatttcacac agaaacagc tatgaccatg attacccaa ggcgcatt
 3601 aacccttact aaagggaaaca aaagctggag ctccaccgcg gtggccgcg ctctagaact
 5 3661 agtggatccc cgggctgca ggaattcgat atcaagctt tegatacgc tgacctcgag
 3721 catcagattg gctattggcc attgcatacg ttgtatccat atcataatat gtacatttat
 3781 attggctcat gtcacacatt accggcatgt tgacattgt tattgactag ttattaatag
 3841 taatcaatta cggggcttact agttcatacg ccatatagg agttccgcgt tacataactt
 3901 acggtaatg gccggcttgg ctgaccggcc aacgaccccc gcccattgac gtaataatg
 10 3961 acgtatgttc ccatagtaac gcaataggg acatttccat gacgtaatg ggtggagtat
 4021 ttacggtaaa ctgcccactt ggcagttacat caagtgtatc atatgtcaag tacgccccct
 4081 attgacgtca atgacggtaa atggccgcg tggcattatg cccagttacat gaccctatgg
 4141 gactttctt cttggcgtt catctacgtt ttatgtatcg ctattaccat ggtgtgcgg
 4201 ttggcagt acatcaatgg gctggatag cggtttgcact cacgggatt tccaagtctt
 15 4261 cacccttgcg acgtcaatgg gagggttggg tggcaccaaa atcaacggga ctttccaaaa
 4321 tgcgttaaca actccggccc attgacgcata atggggcgta ggcgtgtacg gtgggaggtc
 4381 tataaagca gagtcgtttt agtgcacgtt cagatcgctt ggagacgcca tccacgtgt
 4441 ttgcactcc atagaagaca cccggaccga tccagctcc gggccggga acgggtgcatt
 20 4501 ggaacgcgga ttccccgtgc caagagtgc gtaagtaccc cctatagact ctataggcac
 4561 accctttgg ctcttatgc tgctatactg tttttggctt gggcctata cccccccgt
 4621 tccttatgtc atagggtatg gtatagctt gcttataatgt gtgggttattt gaccattatt
 4681 gaccactccc ctatgggtga cgtactttt cttactaaat ccataacatg gcttttgc
 4741 acaactatctt atgggtctt atggcaataa ctctgtctt cagagactga caggactt
 4801 gtatttttac agatggggt cccattttt atttacaaa tcatatatac aacaacgcog
 25 4861 tccccctgca cccgggtttt tattaaacat agcgtggat ctccacgcga atctcggtt
 4921 cgtgtccgg acatgggtca ttcccggtt gggggggggg ttccacatcc gggccctgg
 4981 cccatgcctc cccggcgtca tggcgtctcg gcaacttcc tcccttaaca gtggaggcca
 5041 gacttggca cggcacaatg cccaccacca ccagtgtgc gcaacggcc gtggcggtag
 5101 ggatgtgtc taaaatggat cgtggatgg gggctcgac ggctgacgc gatggaaagac
 5161 ttaaggcgc ggcagaagaa gatgcaggca gctgaggatgt ttttttttgc taagatcg
 5221 agttaactcc cgttgggtt cgttggatgg tggagggttgc ttttttttgc gcaacttcc
 5281 ttgcgtccgc gggggccacc agacataataa gctggacacac taacagactg ttttttttca
 5341 tgggtttttt ctgcgttcc cgtggatca atcatttccat ttttttttgc ttttttttgc
 5401 gtgttttactt atatatctaa atttaatatt ttttttttgc ttttttttgc ttttttttgc
 5461 atgaatttctt caaggatatt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 5521 tcggctgcgc cagagccgaa aggtacccac gtttttttgc ttttttttgc ttttttttgc
 5581 gtaaaggccgg gggggccccc tagatctcc ttttttttgc ttttttttgc ttttttttgc
 5641 gcctggatga gctgggtcccg cccatgcctt gggggggggg ttttttttgc ttttttttgc
 5701 aaaagcaaaa ttgtgtgtt gacaacagac tatgtgcac ccgtggaaagg cagatcc
 40 5761 atctcaagag atgattcaaa aaacacgtt ttttttttgc ttttttttgc ttttttttgc
 5821 gacacagccg tatattactg ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 5881 ccccccattt gggggccaggg aaccctggtc accgttttcc cccatgcctt cccatgcctt
 5941 tcggctttcc cccatgcctt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 6001 tgcgttccgc agactactt cccatgcctt ttttttttgc ttttttttgc ttttttttgc
 45 6061 accagccgcg tgcacaccc ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 6121 aacgtgttca cccatgcctt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 6181 cacaagccca gcaacacccaa ggtggacaaag aaatgttgc gcaatcttgc ttttttttgc
 6241 cacatgcctt cccatgcctt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 6301 ccccaaaaac ccaaggacac cccatgcctt ttttttttgc ttttttttgc ttttttttgc
 6361 gtggacgtga gcaacacccaa cccatgcctt ttttttttgc ttttttttgc ttttttttgc
 6421 gtgcataatgg ccaagacaaa gggggggggg gggggggggg gggggggggg ggggggggg
 6481 agcgttccgc cccatgcctt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 6541 tccaaacaaatg cccatgcctt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 6601 cgagaaccac aggtgttccgc cccatgcctt ttttttttgc ttttttttgc ttttttttgc
 6661 agcctgttccgc gcaacacccaa cccatgcctt ttttttttgc ttttttttgc ttttttttgc
 6721 aatgggcaccc cccatgcctt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 6781 ttcttcttccgc acacacccaa cccatgcctt ttttttttgc ttttttttgc ttttttttgc
 6841 tcatgttccgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 6901 tccatgttccgc acacacccaa cccatgcctt ttttttttgc ttttttttgc ttttttttgc
 6961 gtgttccgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 7021 aaatgtttt ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 7081 gacacacccaa gacacacccaa cccatgcctt ttttttttgc ttttttttgc ttttttttgc
 7141 gccacccatgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 7201 actgttccgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 7261 ggttccgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 7321 gccaacacccaa cccatgcctt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 7381 gcctggaaagg cccatgcctt ttttttttgc ttttttttgc ttttttttgc ttttttttgc
 7441 caaagcaaca acaacacccaa cccatgcctt ttttttttgc ttttttttgc ttttttttgc

7501 tcccacaaaaa gctacagctg ccaggtcactg catgaaggga gcaccgtgga gaagacagtg
 7561 gcccctgcag aatgttcacc gcggaggggag ggaaggccc ttttgaagg gggaggaaac
 7621 ttcgcgcctat gactctcttc gtcccccccg cacggaacac tgatgtcag aggcccctct
 7681 gccattgctg cttctctgc ctttcctctgt cactctgaat gtggcttctt tctactgcc
 5 7741 acagcaagaa ataaaaatctc aacatctaaa tgggttctt gagattttc aagagtcgtt
 7801 aagcacattc cttccccaggc accccctgtc gcaggccagt gccagggcacc aacttggcta
 7861 ctgcgtccca tgagaaaaat ccagttaaat atttccaaa gcaaatggg ttacatatgc
 7921 cctagatctt gattaacagg tggtttgtat tatctgtct ttcgcttcaac ccacattatc
 7981 ccattgcctc cctctcgaggg gggggcccgat accaaatctg ccttatagtg agtcgttata
 10 8041 cgcgcgtca ctggcgtcg ttttacaacg tcgtactgg gaaaacctgt ggttaccca
 8101 acttaatcgc ctgcagcac atccccctt cgcgcgtcg cgttatagcg aagaggcccg
 8161 caccgatcgc cttcccaac agttgcgcag cctgaatggc gaatggaaat tgaagcggtt
 8221 aatattttgt taaaattcgc gttaaatttt tttttaaatca gtcattttt taaccaatag
 8281 gccggaaatcg gcaaatccc ttataaatca aaagaataga ccgagatagg gttgagtgtt
 15 8341 gttcagttt ggaacaagag tccactattt aagaacatgg actccaacgt caaaggcggaa
 8401 aaaaacgtct atcaggcggta tgccggacta ctccggatc atatgacaag atgtgtatcc
 8461 accttaactt aatgtttt accaaaatca ttagggggtt catcgtgt cagggtcaac
 8521 gagaattaaac atccgtcag gaaagctt gatgtatgt tgctaaaaaa ctactcaat
 8581 ggctggttat gcatactcga atacatgcga aaaacctaaa agagcttgc gataaaaaag
 20 8641 gccaattttt tgcttatttac cgcggctttt tatttgcgtt gaaagataaa taaaatagat
 8701 aggtttttt tgaagctaaa tttttttttt cgtaaaaat gccccttgg gttatcaaga
 8761 gggtcattttt atttcgcggta ataaatcatat ttgggtcga aataactaag cacttgcctc
 8821 ctgttactc cctcgatgtt gagggtttaa catgagtc atcgtatgc ggataataat
 8881 acagaaaaac gctaaaccaa ttatccaaat ccagccatcc caaattggta gtgaatgatt
 25 8941 ataaataaca gcaaaacatgt atgggcaat aacacgggtt gcatggtaa ggctcacc
 9001 taatccctgt aaagcacattt gctgtactt ctttgcgtt atagacatca ctccctgt
 9061 tgcaggtaaa gcgatccac caccagccaa taaaattaaa acaggaaaaa ctaaccaacc
 9121 ttcaagatata aacgtttaaa aggcaatgc actactatc gcaataaato cgagcagtag
 9181 tgccgtttt tgcccttattt agtggatattt cttctgcga caaaggcttg gaataactgag
 30 9241 tgtaaaagac caagacccgtt aatggaaaacg caaccatcat gctatttcatc atcacgattt
 9301 ctgtatagc accacacccgt gctggatgg ctatcaatgc gctgaaatata taatcaacaa
 9361 atggcatgt taaaatagtg atgtatccg atcagctttt gttccctta tgagggtt
 9421 attgcgcgtt tggcgtatc atggcgtatag ctgtttctt tttttttttt tttatccgt
 9481 acaattccac acaacatcag agccggaaagc ataaaatgtt aagctgggg tgccataatga
 35 9541 gtgagcttaac tcacattat tgccgttgcgc tcactggccg ctttccatgc gggaaacctgt
 9601 tgcgtccaggc tgcatatgtt aatccggccaa cgcgcggggg gagggggtt gcttatttt
 9661 cgcttcccg ctccctcgct cactgactcg ctgcgtcg tggcgtcg gggcgagcgc
 9721 gtatccatgc actcaaaaggc ggttaatccatgc ttatccatgc aatcaggggta taacgcagg
 9781 aagaacatgtt gaccaaaaggcc caagcaaaag gccaggaaacc gtaaaaaaggc cgcgttgc
 40 9841 gctttttcc ataggctccg ccccccgtt gacccatcaca aaaatcgacg ctcaagttag
 9901 aggtggcga acccgacagg actataaaaga taccaggcg tttccctgg aagetccctc
 9961 gtgcgttccctc ctgttccgac cctgcgttcc accggatacc tttccctgg tttccctgg
 10021 ggaaggctgg cgttttctca tagtgcacgc tgtaggtatc tcagttgggt tgtaggtgtt
 10081 cgcttcaaggc tggtgtgttgc gcaacggacc cccgttccatgc cggccgttgc
 45 10141 ggttaatctc gtttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 10201 actggtaaca ggttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 10261 tggcttaact acggctacac tagaaggaca gtttttttttccatgc ttttttttttccatgc
 10321 gttaccttcg gaaaaagatgt tgtagtctc tgatccggca aacaaaccac cgttggtagc
 10381 ggtgtttttt ttgtttccatgc gcaacggacc acggccggcc aaaaaggatc ttttttttttccatgc
 10441 ccttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 50 10501 ttgttcatgtt gattatccaa aaggatcttccatgc ttttttttttccatgc ttttttttttccatgc
 10561 tttaaatccaa ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 10621 agtggccac ctatccatgc gtttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 10681 gtcgtttaga taactacatgc acggggggcc ttttttttttccatgc ttttttttttccatgc
 55 10741 ccgcggccac cacgttccatgc gtttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 10801 gcccggccca gaaatggcc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 10861 cggggggcc gaaatggcc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 10921 acggccatcg ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 60 10981 cgatccatgc gtttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11041 cctccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11101 ctgcataatt ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11161 tcaaccaatgtt ctttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11221 atacgggata atacccgcgc acatagcaga acttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11281 ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11341 actcgccac ccaactgttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11401 aaaaacggaa gggaaaaatccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11461 ctcatactct ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc
 11521 ggatacatat ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc ttttttttttccatgc

3901 ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac
 3961 gccaagcgcg caattaaccc tcactaaagg gaacaaaagc tggagctcca cgcgcgtggc
 4021 ggccgccta gaactagtgg atccccggg ctgcagaaat tcgatatacgat
 4081 accgcgtgacc tcgagctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag
 5 4141 cttgacctga tacctgattt tcttcaaaa gggaaaacaa cacaatccca caaaacagct
 4201 cagagagaaa ccatcactga tggctacagc accaaggat gcaatggcaa tccatcgac
 4261 attcatctgt gacctgagca aaatgattt tctctccatg aatggttgct tcttccctc
 4321 atgaaaaggc aatttccaca ctcacaat ggcacaaaaga caaacagaga acaattaatg
 10 4381 tgctccttc taatgtcaa attgttagtgg caaagaggag aaaaaatct caagtctga
 4441 ttaggtttt gtaggtggat aagaggctt gctgttgtag ctcacccggaa cttcatatcc
 4501 ttttgataa aaagtgcctt tataacttc aggtctccga gtcttatttc atgagactgt
 4561 tggtttaggg acagaccac aatgaaaatgc ctggcatagg aaaggcagc agagccttag
 4621 ctgacccccc ctgggacaa gcattgtcaa acaatgtgtg acaaaaactat ttgtactgct
 4681 ttgcacacgt gtgctggca gggcgttcca ttgcacacta tcccaggtaa ccttccaaact
 15 4741 gcaagaagat ttgtgcctac tctctctaga aagcttcgtc agactgacat gcatttcata
 4801 ggttagagata acatttactg ggaaggcacat ctatcatc aaaaaggcagg caagatttc
 4861 agactttttt atgtggctgaa atagaaccaa aagcgttaat taaaacaaa atgaaacaaa
 4921 aaaaatcagt tgataccctgt ggttagaca tccagcaaaa aaatattt tgcactacca
 4981 tcttgtctta agtccctcaga ctttagcaagg agaatgttaga tttccacagt atatatgttt
 20 5041 tcacaaaagg aaggagagaa aaaaaagaaaa atggcaactgca ctaaacttca gctagtggt
 5101 taggaaaagta attctgccta acagagattt cagtgtatctc tatgtatgtc ctgaagaatt
 5161 atgtgtactt tttttcccccc atttttaaat caaacagtgc tttacagagg togaatgg
 5221 ttctttaactg ttgtcaattt otatattttt aatacacaac aataacttct ataaactgaaa
 5281 tattttgtt atttgtatattt atttgttcc ctcgaacat gaacacttcc ccaactgtaat
 25 5341 ttcacaattt ctctgtcatc tgccaggcca ttaagtattt catgaaagat ctttgaggaa
 5401 cactgcaagt tcataatcata aacacattt aatttggat ttgggttgca ttgtatggag
 5461 ctatgtttt ctgtatccctc agaaaaaaag tttgttataa agcattcaca cccataaaaaa
 5521 gatagattta aatattccag ctataggaaa gaaagtgcgt ctgccttca ctctagtctc
 5581 agtgggtcc tcacatgcg tgccttcattt tttgtcaaga aataatagg
 5641 tcacgttcttgc ttctcaattt tgcctcgcc agcatgtc agatgcacgt ttagataca
 5701 agaaggatca aatggaaacag aatttgcattt tttttccat tctcaaggat cccatatttt
 5761 taactaataa ttgctaattt tttttccat ctctcaaggat tttttccat tttttccat
 5821 taaagatccc attatctggt tgtaactgaa gctcaatggaa acatggacaa tttttccat
 5881 tcttctccat catccaaacag tccgtatggaa ttagcagaac aggcagaaaa cacattgtt
 5941 cccagaattt aaaaactaataa tttgtctccat attcaatcca aataggacat attgaaacta
 6001 aaatctaacc caatccattt aatgttattt tatgtgtca aaggctaaac ttctgaagg
 6061 aacctgtggg tgggtcacaat ttcaggctat atatttccca gggctcagcc agtggatcaa
 6121 tcattcatct cgtgacttct tgcgttggg tttttccat tataataaa tttatattt
 6181 cgttattaa aattttatattt ttttcgacca tgaatttcc aaggatattt ttcttcgtgt
 6241 tcgctttggt tctggctttt tcaacagttt cggctgcgc agagccgaat ggtacccagg
 6301 tgcagctgca ggagtgggggg ggaggcttgg taaagccggg ggggtccctt agagtctcc
 6361 gtgcagccctc tggattcaat ttcagaaacg cctggatggag ctgggtccgc caggctccag
 6421 ggaaggggct ggagtgggtc ggccgttataa aagcaaaaat tgatgggtggg acaacagact
 6481 atgcgtcacc cgtgaaaggc agatgtccca tctcaagaga tgatccaaaa aacacgttat
 6541 atctgcataat gatagccctg aagccgggg acacaggcgtt atattactgt accacccgg
 6601 ttatgataac attggggggaa gttatccctt ccccaatttgg gggccaggaa accctggtca
 6661 ccgttctccat agccctccacc aaggggccat cggcttccctt cctggcacccttccat
 6721 gcacccctgg gggccacagcg gcccctgggtt gcctggtaa ggactacttcc cccgaacccgg
 6781 tgacgggtgc tggtaactca ggcgccttgc ccagccgggt gcacacccccc cccgctgtcc
 6841 tacatccctc aggactctac ttcccttgcac acgtgggtac cgtgcccctcc accagcttgg
 6901 gcacccagac ctacatctgc aacgtgaatc acaagcccgac caacaccaag gtggacaaga
 6961 aagtggaccc caaatccatcc accgtggccca acatggccca gcaacccccc
 7021 tcctgggggg accgtcagtc ttcccttccat ccccaaaaacc caaggacacc ctatgatct
 7081 cccggacccc tgaggtcaca tgcgtgggg tggacgttgc ccacggaccc cctgggggtca
 7141 agttcaactg gtacgtggac ggcgtggagg tgcataatgc caagacaaat cccgggggg
 7201 agcagtacaa cagcacgtac cgtgtggtca gcgtccctcact cgtccctgc caggacttgg
 7261 tgaatggcaaa ggagtacaag tgcagggttcc ccaacaaatg cctccctccccc cccatcgaga
 7321 aaaccatctc caaaggccaaa gggcagcccc gagaacccca ggtgtacacc ctggccccc
 7381 cccggatga gtcgaccaag aaccggatca gcctggacttgc cctggtccaa ggccttctatc
 7441 ccagcgacat cggccgtggag tggggagggc atggggcaggcc ggagaacaaac tacaagacca
 7501 cgcctccctcgt gtcggactcc gacggcttgc tctcttgcata cagaacgtcc accgtggaca
 7561 agacgggggt gcacgggggg aacgtcttctt catgtccgt gatgttgc gctctgcaca
 7621 accactacac gcagaagggc ctctccctgt ctccgggttca agcggccaggcc cccaaatctt
 7681 cctatggact gacacagccca ccctccgggttca cagtttttttcc accacaaacg gccaggatca
 7741 cctgtctgg agatgttccat ccagaaaaat atgtttatttgc ttttttttttcc accacccgg
 7801 aggccccctgt ggtgggttcaat tatggggacca gcaaaacggacc ctcggggatcc cctggagat
 7861 tctctggctc cagtcaggg acaatggccca ctttgactat cagtggggcc caggtggaaag
 7921 atgaagggtga ctactactgt tactcaactg acagcgttggg tttatcatagg gaggtgttca

5	7981	gccccggggac	caagctgacc	gtccttaggtc	agccccaggc	tgccccctcg	gtcaactctgt
	8041	tcccccaccc	ctctgaggag	cttcaagccca	acaaggccac	actgggtgt	ctcataaagtg
	8101	actccctaccc	gggagccgtg	acagtggcct	ggaaggcaga	tagcagcccc	gtcaaggcgg
	8161	gagttggagac	caccacaccc	tccaaacaaa	gcaacaacaa	gtacgcggcc	agcagctacc
	8221	tgagccgtac	gcttgagcag	tggaaatccc	acaaaagcta	cagctccag	gtcacgcatg
	8281	aaggggacac	cgtggagaag	acagtggccc	ctgcagatgt	ttcaccegg	aggggggaa
10.	8341	ggggcccttt	tgaaggggga	ggaaatctcg	cgccatgact	ctctctcg	ccccccgacg
	8401	gaacactgtat	gtgcagaggg	ccctctgcca	ttgtgtctt	ctctgcctt	cctctgtact
	8461	ctgaatgtgg	cttcttgc	actgcccacag	caagaataaa	aatctcaacc	tctaaatggg
	8521	tttctgtgaga	tttttcaaga	gtcgttaagc	acattccctt	cccagcaccc	tttgctcgag
	8581	gccaggccca	ggcaccaact	tggctactgc	tgcccatgag	agaaaatccag	ttcaatattt
15	8641	tccaaagacca	aatggattac	atatgcocca	gatctgtatt	aacagggtt	ttgtattatc
	8701	tgtgtttcg	cttcacccac	attatccat	tgcccttccct	cgaggggggg	ccccgttacc
	8761	aattccgtcc	atagtggat	gtattacggc	cgctcaatgg	cgctgtttt	acaacgtcg
	8821	gactggggaa	accctggcg	tacccaaactt	aatgccttg	cagcacatcc	ccctttcgcc
	8881	agctggcgta	atagcgaaga	ggcccgccacc	gatgcctt	cccaacagg	gcccagctg
	8941	aatggcgaaat	ggaaatttgc	agcgttaata	ttttgttaaa	attcgcgtt	aattttgtt
	9001	aaatcagctc	attttttaac	caataggccg	aaatcgccaa	aatccctat	aatcaaaaag
20	9061	aatagaccga	gatagggtt	agtgtgtt	cagtgtggaa	caagagtcca	ctattaaaga
	9121	acgtggactc	caacgtcaaa	gggcggaaaaa	ccgtctatca	gggcgtatgc	ccactactcc
	9181	gggatcatat	gacaagatgt	gtatccac	taacttaatg	attttacca	aaatcattag
	9241	gggatttcata	agtgtctcagg	gtcaacgaga	attaacatc	cgtcaggaaa	gctttagatc
	9301	atgtatgtgt	taaaaactt	ctcaatggct	ggttatgc	atcgcataac	atgcggaaaa
	9361	cctaaaaagag	cttgcgata	aaaaaggcca	atttattgt	atttaccgcg	gtttttattt
	9421	gagcttgaaa	gataaataaa	atagataggt	tttatttgaa	gtctaaatett	ttttatcgta
	9481	aaaaatgccc	tcttgggtt	tcaagagggt	cattatattt	cgccggataaa	catcatattg
	9541	tgacgaaata	actaagca	tgtctctt	ttactccct	gagcttgcgg	ggttaacatg
	9601	aagggtcata	atagcaggat	ataatacag	taaaacgcta	aaccaataat	ccaaatccag
	9661	ccatcccaaa	tttgcgtat	atgattataa	ataacagcaa	acagataatgg	gccaataaaca
30	9721	ccgggttgcat	ttgtgttgc	caccaataat	ccctgttaaag	caccttgcgt	atgactcttt
	9781	gttggatag	acatcactcc	ctgtatgc	ggtaaagcga	ttccaccacc	agccaataaaa
	9841	attaaaaacag	ggaaaactaa	ccaaaccttca	gatataaaacg	ctaaaaaggc	aatgtcacta
	9901	ctatctgca	taaatccgg	cagtaactgc	gttttttgc	ccatattagt	gttatttctt
	9961	ctgccccaaa	ggcttggat	actgagtgt	aaagaccaag	accctgtat	aaaagccaaac
35	10021	catcatgtca	ttcatatca	cgatttctt	aatagacca	caccctgt	gatgggtat
	10081	caatcgctg	aaaaataata	caacaaatgg	catcgataaa	taatgtat	ataccgtatca
	10141	gtttttgttc	ccttttagt	gggttaattt	cgccgttgc	gtatcatat	tcatagctgt
	10201	ttctgtgtgt	aaattgttat	ccgctcaca	ttccacacaa	catacgagg	ggaagcataaa
	10261	agtgtaaagc	ctgggggtgc	taatgtgt	gctaactc	attaattgc	ttgcgtc
	10321	tgcccgctt	ccagtcggga	aacctgtcg	gcccagctgc	ttaatgaatc	ggcaacacg
	10381	cggggagagg	cggtttgcgt	atttgggcgt	cttccgtt	ctcgcatact	gactcgtcgt
	10441	gtctggctgt	tcggctgcgg	cgagcgtat	cagtcactc	aaaggcgtt	atacgggtat
	10501	ccacagaatc	aggggataac	cgaggaaaga	acatgtgagc	aaaaggccag	caaaggccaa
	10561	gaaacccgtaa	aaaggcccg	ttgtgttgc	tttccatag	gtccggcccc	cctgcacgagc
	10621	atcacaaaaaa	tcgacgtca	agtcaagggt	ggcgaaaacc	gacaggacta	taaagatacc
	10681	aggcgtttcc	cccttggaa	tcctctgt	gtctcttgc	tccgaccct	cegcttaccg
	10741	gatacctgtc	cgcccttctc	ccttcgggaa	gcttggcg	ttctcatagc	tcacgtgt
	10801	ggtatctcg	ttcgggtgt	gtcttgcgt	ccaaagctgg	ctgtgtgcac	gaaccccccgg
	10861	ttcagccca	ccgtgtgc	ttatccggta	actatcg	tgagtccaa	coggtaaagac
	10921	acgacttatac	gccaatcgca	gcagccact	gtaaaggat	tagcagacg	aggtatgtag
	10981	gccccgtac	agatctt	aatgtgttgc	cttaactacgg	tcatactaga	agacagat
	11041	tttgttatct	cgcttgcgt	aaggcgtt	ccttcggaaa	aagagtttgt	agctctgtat
	11101	ccggcaaaca	aaccaccgt	ggtagegggt	tttttttgt	ttgcacgc	cagattacgc
	11161	gcagaaaaaa	aggatctaa	gaagatctt	tgatctttc	tacgggtt	gacgctcgt
	11221	gaaacggaaa	ctcacgtt	gggattttgg	tcatgagatt	atcaaaaagg	atcttaccc
	11281	agatctttt	aaattaaaaa	tgaatttt	aatcaatctt	aagtatata	gagtaaactt
	11341	ggttgcac	ttaccaatgc	ttatctgt	aggccacat	ctcagcgatc	tgtctatcc
	11401	tttcatccat	agttgcgt	ctccccgt	tgtatgtat	tacgatacgg	gggggttac
	11461	catctggccc	cagtcgtca	atgatccgc	gagacccacg	tcaccggct	ccagatttt
	11521	cagcaataaa	ccagccagcc	ggaaggccgg	agcgcagaag	tggctctgca	actttatccg
	11581	cctccatcca	gtcttataat	tgttgcggg	aaatctgt	aagtgttgc	ccagttataa
	11641	gtttgcgc	cggtttgc	attgtat	gatctgtgt	gtcacgc	tcgttttgt
	11701	tggcttcat	cagtcgttgc	tccaaacgt	caaggcgt	tacatgtat	cccatgtt
	11761	gcaaaaaaa	ggttagtgc	ttcggcttc	cgatctgtt	cagaagta	ttggccgcag
	11821	tgttatct	catgttgc	gcagcact	ataattctt	tactgtcat	ccatccgtaa
	11881	gatctttt	tgtgtactgt	gtagtact	ccaaatgtt	ctgagaatag	tgtatgcgg
	11941	gaccgagtt	cttgcgtcc	gctgtcaat	gggtatcc	cgccgcacat	agcagaactt
60	12001	taaaatgtgt	catcatttgc	aaacgtt	tttttttttt	actctcaagg	atcttaccc
65							

5	7741	tcacccgtc	tggagatgc	ttggccagaaa	aatatgttta	ttggtaccag	cagaagtca
	7801	gccaggcccc	tgtgggtgc	atctatgagg	acagcaaaacg	accctccggg	atccctcgaga
	7861	gattctctgg	ctccagtc	gggacaatgg	ccacccctgac	tatcagtgg	gcccaggtgg
	7921	aagatgaagg	tgactactac	tgttactcaa	ctgacagcag	tggttatcat	agggaggtgt
	7981	tcagccggg	gaccaagctg	acccgtctag	gtcagcccaa	ggctgcccc	tcggtaactc
10	8041	tgttccacc	ctccctgtg	gagcttcag	ccacaaggc	cacactgg	tgtctcataa
	8101	gtgactccct	cccgggagac	gtgcacgtgg	cctggagggc	agatagcgc	cccgtaagg
	8161	cggggagtgg	gaccacaca	ccctccaaac	aaagcaacaa	caagtacgc	gccagacgt
	8221	acctgagct	gacgcttgc	cagtggaaagt	cccacaaaag	ctacagctc	caggtaacgc
	8281	atgaagggg	caccgtgg	aagacagtgg	cccccgtcaga	atgttcac	eggaggagg
	8341	gaaggggccct	ttttgaagg	ggaggaaact	tcgcgcctat	actccctctg	tgccccccgc
	8401	acggaaacact	gatgtgcaga	gggcctctg	ccattgtctc	ttccctctgc	ttccctctgc
	8461	actctgaatg	tggtttctt	gctactgc	cagcaaaaaa	taaaatctc	acatctaaat
15	8521	gggttctgt	agattttca	agatcgta	acacatcc	tccccaga	cccccttgcgt
	8581	caggccatgt	ccaggccacca	acttggctac	tgctgccccat	gagagaaatc	cagttcaata
	8641	ttttccaaag	caaaatggat	tacatatgc	ctagatctc	attaacagg	gttttgatt
	8701	atctgtgc	tcgcttcacc	cacattatcc	cattgcctcc	cctcgaggg	gggccccgt
	8761	cccaattcgc	cctatagta	gtcgattac	gcccgcctac	tggccgtcgt	tttacaacgt
20	8821	cgtgactgg	aaaacccctgg	cgttacccaa	cttaatcgcc	ttgcagcaca	tcccccttcc
	8881	cccgatgtgc	gtatagcga	agaggccgc	acccatgcgc	cttcccaaca	gttgcgacgc
	8941	ctgaatggcg	aatggaaat	gtaaagctt	atattttgtt	aaaattcgc	ttaaattttt
	9001	gttaaatcag	ctatctttt	aaccaatagg	ccggaaatcg	caaaatcc	tataaatcaa
	9061	aagaatagac	cgagataggg	tttagtgc	ttccagtttgc	gaacaagagt	ccactattaa
25	9121	agaacgtgg	ctccaaacgtc	aaaggcgaa	aaaccgtcta	tcagggcgat	ggcccactac
	9181	tccgggatca	tatgacaaga	tgtgtatcc	ccttaactta	atgattttta	ccaaaatcat
	9241	taggggattc	atcagtgc	agggtcaacg	agaattaaca	ttccgtcagg	aaagcttgc
	9301	atgtatgtat	gcttaaaaac	ttactcaat	gtcggtttag	catatcgca	tacatgcga
	9361	aaacccataaa	gagtttgc	aaaaaaagg	ccaaattttt	gtatatttc	gcccgtttt
30	9421	attgagctt	aaagataat	aaaatagata	ggtttttttt	gaagcttaat	tttcttttate
	9481	gtaaaaaaatg	cccttgggg	ttatcaagag	ggtcattata	tttcgcggaa	taacatcatt
	9541	tggtgcgaa	ataactaaagc	acttgc	tttttactcc	cctgagctg	aggggttaac
	9601	atgaaggtca	tcgatagc	gataataata	cagtaaaaacg	ctaaaccaat	aatccaaatc
	9661	cagccatccc	aaatttggtag	tgaatgatta	taaataacag	caaacagtaa	tgggccaata
35	9721	acacccgttgc	catgttgc	gtcacaat	aaatccctgt	cgcacccat	ctgatgc
	9781	tttttttgc	tagacatcc	ttccctgt	gcaggtaaag	cgatccccc	accagccat
	9841	aaaataaaaa	caggaaaaac	taaccacac	tcgatataaa	acgctaaaaa	ggcaaatgc
	9901	ctactatctg	caataaaatcc	gagcgtact	ggccgtttttt	cccccattt	gtggctattc
	9961	ttccgtccac	aaaggcttgc	aatactgagt	gtaaaagacc	aagaccctgt	ataaaaagcc
40	10021	aaccatcatg	ctattcatc	tcacgat	tgtatagca	ccacaccctg	ctggatttgc
	10081	tatcaatgc	ctgaaataat	aatcaacaa	tgcatcg	aaataagt	tgtataccga
	10141	tcagtttttgc	tttcccttgc	tgagggttta	tttgcgc	ggcgtatca	tggtcatagc
	10201	tgttttctgt	tgaaatgt	tatccgtca	caattccaca	caacatagca	ggccgaagca
	10261	taaaaatgtaa	agccctgggt	gcctaata	tgagctaa	cacattaaat	ggcgttgc
45	10321	caetgcccc	tttccagtc	ggaaacctgt	cgtgcac	gcattaaatg	atgcgcac
	10381	gcgcggggag	aggcggttgc	cgtattgggc	gtcttccgc	ttcctcgctc	actgactcgc
	10441	tgcgtcggt	cgttccgt	cgccgacgg	tatcgtc	ctcaaaaggcg	gtataacgg
	10501	tatccacaga	atcaggggat	aacgcaggaa	agaacatgt	agcaaaaaggc	cagcaaaaagg
	10561	ccaggaaccc	taaaaaggcc	gcgttgc	cgtttttcc	taggtcccg	ccccctgcac
50	10621	agcatcaca	aaatcgac	tcaatgc	ggtggc	ccgcacagg	ctataaaatg
	10681	accaggcg	tcccccgttgc	agctccctgc	tgccgtctcc	gtcttccat	agtcacgc
	10741	ccggatacc	gtccgc	ctcccttc	gaagcgtgc	gggtctgtgt	cacgaacccc
	10801	gttaggtatc	cagttccgt	taggtcg	gtcccaacgt	tcttgcgacc	taccccgtaa
	10861	ccgttcagcc	cgaccgc	gccttata	gtaaatc	gttttgcgat	ttctacgggg
55	10921	gacacgactt	atcgccact	gcagcagcc	ctggtaa	gattagcaga	gcgaggat
	10981	taggcgttgc	tacagatgc	ttgaatgtt	ggcctaacta	cggtctacact	agaaggacag
	11041	tatccatgt	ctggcgtct	ctgatgc	ttacccctgg	aaaaagatgt	ggtagctt
	11101	gatccgc	acaaaccacc	gtgttgc	gtggttttt	tgtttgc	cacgcagat
	11161	ccgcgcagaa	aaaaggat	caagaagatc	cttgc	ttctacgggg	tgcacgtc
60	11221	agtggaaac	aaactc	taagggttt	tgttgc	attatcaaa	aggatctca
	11281	cctagatcc	ttttaat	aatgaatgtt	ttaatcaat	ctaaagtata	tatgat
	11341	cttggctc	cgat	tgcttac	gtgaggcacc	tatctcagcg	atctgtctat
	11401	tgcgttc	catagttgc	tgact	cccccc	aactacgata	cgggagggt
	11461	tatccatctgg	ccccagtc	gcaat	gac	acgtc	acgtccagatt
	11521	tatcagcaat	aaacc	ggcggaa	ccggacgg	aagtggct	gcaactt
65	11581	ccgcctccat	cgttctt	atttgt	ggggactgt	agtaatgt	tgcggat
	11641	atagtttgc	caacgttgc	gcatt	ctgc	gtgtca	tcgtcg
	11701	gtatggctt	attcagtc	ggttccaa	gtcaaggc	attatcat	ttccccatgt
	11761	tgtgc	aaaaaa	agcggttgc	tccttcgg	tgtcaga	aagttggcc

11581 gctcttgcac ggcgtcaata cgggataata cccgcacaca tagcagaact ttaaaagtgc
11641 tcatcatgtgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgtttagat
11701 ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca
11761 gcgttctgg gtgagcaaaa acagggaggc aaaatgcgcg aaaaaaaggga ataagggcga
11821 cacggaaatgg ttgaataactc atactcttc ttttcaata ttattgaagc atttatcagg
11881 gttattgtct catgagcgga tacatatgtt aatgtatTTTtta gaaaataaaa caaatagggg
11941 ttccgcgcac atttccccga aaagtgcac

5